ERIC Document Resume | **See Instruction Sheet** | DATE OF RESUME | DATE RECEIVED

A. PUBLICATION DATA

1. ERIC NUMBER
ED001250

2. ORIGINATOR CONTROL NO(S).
NA

3. IDENTIFICATION NO.
CR 1217

DO NOT WRITE IN SPACE BELOW

MICROFILM COPY AVAILABLE
YES ☒ NO ☐

COPYRIGHT GRANTED
YES ☐ NO ☒

COPYRIGHT RELEASE FILED
YES ☐ NO ☒

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

4. AUTHOR(S):
Wallen, Norman E. and Wodtke, Kenneth H.

5. TITLE OF DOCUMENT ABSTRACTED:
Relationships Between Teacher Characteristics and Student Behavior - Part I

6. PUBLICATION TITLE:
N.A.

7. EDITOR(S):
N.A.

8. PLACE OF ORIGIN
ORG. Dept. of Educational Psychology
ST. University of Utah
CITY Salt Lake City, Utah
TELE. EXT.
INST/LOC CODE

9. PUBLISHER OR SOURCE
Cooperative Research Prog., Dept. of HEW, Office of Education - Wash., D.C.

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
N.A.	N.A.	Nov. 63	114

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR
experimental research

16. INDEXING TERMS (6-12)

Classroom dynamics	Student gain	Student responses	Teacher dominance
Elementary educa.	Student-teacher interaction		Teacher effectiveness
Student behavior	Teacher characteristics		

17. ABSTRACT (MAX. 250 WORDS)

Pupil data, including academic, emotional and attitudinal characteristics, was correlated with teacher characteristics. There was a clear indication that relationships did not take the same form across the five grade levels of the study (grades one through five). The general impression resulting was that the typical first grader, being somewhat unsure of himself and new to the situation, is both more comfortable and achieves better given a situation which is quite structured, controlled by the teacher. At the same time the teacher should be rather supportive and encouraging but should deemphasize overt affection. By about the second grade this pattern tends to change and by the upper grades the general desirability of encouragement still exists but the effect of control shifts such that a greater degree of permissiveness has the more desirable effects in terms of both liking for school and achievement gain. As a four year study 65 teachers and their students (grades one through five) were involved. Analysis of the data consisted of correlational and factorial analyses.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION

20. DIVISION OR DEPT.

21. CONTRACT NO.

PROJECT NO.

22. TOTAL COST OF PROJECT

23. DURATION OR PROJECT

U.S. Department of Health, Education, and Welfare
Office of Education

2

1217

ED001250

Relationships Between Teacher Characteristics and Student Behavior — Part I

By

NORMAN E. WALLEN AND KENNETH H. WODTKE

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Final Report **November, 1963**

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION CONTRACT NO. 2-10-013

University of Utah

Department of Educational Psychology

RELATIONSHIPS BETWEEN TEACHER CHARACTERISTICS AND STUDENT BEHAVIOR--PART I

by

Norman E. Wallen and Kenneth H. Wodtke

Final Report November, 1963

U. S. Department of Health, Education, and Welfare
Office of Education Contract No. 1217

University of Utah

Department of Educational Psychology

RELATIONSHIPS BETWEEN TEACHER CHARACTERISTICS AND STUDENT BEHAVIOR--PART I

Cooperative Research Project No. 1217

by

Norman E. Wallen and Kenneth H. Wodtke

University of Utah

Department of Educational Psychology

1963

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ACKNOWLEDGEMENTS

Sincere appreciation is expressed to the sixty-five teachers and their pupils whose cooperation made possible the collection of the data for this study. The development of good working relations with the teachers was facilitated by the following principals: Mr. R. J. Bjorkland, Mrs. Maurine McDonald, Mrs. Constance Pickell, Mr. Arnold Lund, Mr. George Miller, Mr. Matthew Galt, and Mr. Russell Goodman. Support at the district level was provided by Superintendent Harold G. Holt of Davis County and Mr. Arthur E. Arnesen, Administrative Assistant, Salt Lake City School District.

Important contributions to the project were rendered by two research assistants, Mr. Joseph L. Daly and Mr. Robert E. Jester.

Appreciation is expressed to the Western Data Processing Center at U.C.L.A. for services rendered in the processing of the data and to Mr. Ian E. Reid for his help in this regard.

Finally, appreciation is expressed to Carolyn Bennion for her skills as typist, accountant, secretary, and reducer of blood pressure.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

CHAPTER PAGE

I. THE PROBLEM 1.01

- Phase I 1.02
- Phase II 1.03

II. REVIEW OF RELATED LITERATURE 2.01

- The Measurement of Classroom Behavior 2.01
- The Prediction of Teacher Classroom Behavior. 2.03
- The Effects of Teacher Characteristics and Teacher Classroom Behavior on Pupil Behavior and Achievement. 2.05

III. PHASE I: PROCEDURE 3.01

- Sample of Teachers 3.01
- Teacher Measures 3.01
- Pupil Measures 3.05
- Administration and Scoring of Measuring Instruments 3.09
- Data Analysis 3.12

IV. PHASE I: RESULTS 3.01

- Analysis of Combined Grades Matrix. 4.02
- Analysis of Teacher Matrices by Grade 4.05
- Analysis of Correlations Within Grade 1 . . . . 4.12
- Relationships Between Observer Perceptions and Pupil Perceptions 4.13
- Results of Factorial Analyses 4.16
- Correlations Among Gains Achieved by Teachers in the Three Achievement Areas 4.24

V. PHASE II: PROCEDURE 5.01

- Selection of High and Low Controlling Teachers. 5.01
- Pupil Measures 5.02

VI. PHASE II: RESULTS 6.01

- The Effect of Teacher Controlling Behavior in the Classroom On Pupil Gains in Creativity or Measured by the Torrance Tests 6.01
- Relationships Between Pupil Classrooom Behavior, Pupil Creativity and Teacher Controlling Behavior . . 6.10

CHAPTER PAGE

The Test-Retest Reliability of the Torrance Creativity Battery 6.22
Relationships Between Measures of Creativity and Intelligence. 6.26

VII. SUMMARY AND CONCLUSIONS 7.01

REFERENCES R.01

APPENDIXES

A. SCORING ADDITIONS TO RUSSELL SAGE TEST OF SOCIAL RELATIONS A.01

B. BETAS, MEANS, STANDARD DEVIATIONS, AND MULTIPLE r's FOR ACHIEVEMENT GAIN SCORES--BY GRADE B.01

C. OBSERVATION RECORD AND GLOSSARY OF RATINGS FOR PUPIL BEHAVIOR C.01

D. PUPIL QUESTIONNAIRE--KEYED D.01

LIST OF TABLES

TABLE PAGE

3.01 COMPARISON OF 'PUPIL CHARACTERISTIC' GROUPS WITHIN EACH GRADE--ANALYSIS OF VARIANCE I 3.15

4.01 CORRELATIONS BETWEEN FACTOR 1 AND SELECTED OSCAR MEASURES ACROSS GRADES 4.06

4.02 SUMMARY OF THE MOST PROMISING RELATIONSHIPS BETWEEN TEACHER MEASURES AND PUPIL MEASURES. 4.11

4.03 INTERCORRELATIONS AMONG MEASURES FOR GRADE 1 . . . . . . 4.14

4.04 INTERCORRELATIONS AMONG MEASURES OF TEACHER BEHAVIOR ACROSS THE FIVE GRADES. 4.15

4.05 RELATIONSHIPS BETWEEN OBSERVER PERCEPTIONS AND PUPIL PERCEPTIONS OF TEACHERS ACROSS GRADES. 4.18

4.06 ANALYSIS OF VARIANCE I--READING VOCABULARY GAIN. . . . 4.19

4.07 ANALYSIS OF VARIANCE I--READING VOCABULARY GAIN--GRADE 1 . . . . 4.20

4.08 ANALYSIS OF VARIANCE II--READING VOCABULARY GAIN--GRADE 1 . . . . 4.21

4.09 ANALYSIS OF VARIANCE II--READING VOCABULARY GAIN--GRADE 2 . . . . 4.22

4.10 ANALYSIS OF VARIANCE II--ARITHMETIC GAIN--GRADE 1. . . 4.23

4.11 CORRELATIONS AMONG ACHIEVEMENT GAINS (BY CLASS) ACROSS THE FIVE GRADES 4.25

4.12 CORRELATIONS AMONG ACHIEVEMENT GAINS (BY INDIVIDUAL) ACROSS THE FIVE GRADES 4.26

4.13 ESTIMATED RELIABILITIES OF CLASS GAIN SCORES BY GRADE AND CONTENT AREA 4.27

6.01 MEANS AND STANDARD DEVIATIONS OF THE REGRESSED CREATIVITY TEST GAIN SCORES FOR EACH TEACHER 6.02

6.02 ANALYSIS OF VARIANCE OF TOTAL VERBAL CREATIVITY GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE 4 . . 6.04

6.03 ANALYSIS OF VARIANCE OF VERBAL FLEXIBILITY GAINS FOR HIGH AND LOW CONTRLLING TEACHERS IN GRADE 4 . . . . 6.06

TABLE PAGE

6.04 ANALYSIS OF VARIANCE OF VERBAL FLUENCY GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE 4 . . . . 6.07

6.05 ANALYSIS OF VARIANCE OF NON-VERBAL ELABORATION GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE 4. . . . 6.09

6.06 THE RELIABILITY OF NINE CATEGORIES OF PUPIL BEHAVIOR AND THE RATINGS OF PUPIL BEHAVIOR 6.12

6.07 RELATIONSHIPS BETWEEN PUPIL CLASSROOM BEHAVIOR AND MEASURES OF PUPIL INTELLIGENCE AND CREATIVITY....... 6.15

6.08 ANALYSIS OF VARIANCE OF THE NON-CREATIVITY BEHAVIOR IN THE CLASSROOM FOR HIGH AND LOW CREATIVE PUPILS OF HIGH AND LOW CONTROLLING TEACHERS 6.16

6.09 ANALYSIS OF VARIANCE ON THE SELF-INITIATED PUPIL TALK IN THE CLASSROOM FOR HIGH AND LOW CREATIVE PUPILS OF HIGH AND LOW CONTROLLING TEACHERS 6.19

6.10 ANALYSIS OF VARIANCE OF THE OBSERVER RATING APATHETIC-ALERT FOR HIGH AND LOW CREATIVE PUPILS OF HIGH AND LOW CONTROLLING TEACHERS 6.21

6.11 NON-VERBAL TEST-RETEST RELIABILITIES FOR GRADES 2 THROUGH 5 . . . . 6.24

6.12 VERBAL CREATIVITY TEST-RETEST RELIABILITIES FOR GRADES 4 AND 5 . . . . 6.25

6.13 RELATIONSHIPS BETWEEN INTELLIGENCE AND CREATIVITY IN GRADES 2 THROUGH 5. . . . . 6.27

CHAPTER I

THE PROBLEM

It is increasingly apparent that educational research must begin to establish relationships between certain environmental events to which students are exposed in school and consequent behavior of the students. A prerequisite to the study of such questions is the specification of differences in the environmental situations themselves. One major category of environmental differences is that of differences among teachers. A number of recent developments have suggested meaningful dimensions of teacher difference and have begun to spell out in some detail techniques to be used in measuring these difference variables.

One outcome of a study conducted under Project No. 444 (8029) of the U. S. Office of Education (Travers, Wallen, Reid, and Wodtke, 1961) was the specification of a number of dimensions of teacher characteristics based on analyses of verbatim classroom verbal behavior, observed classroom behavior, expressed attitudes, and response to projective devices. As a result of this study, one hundred and eighteen elementary school teachers (grades K-6) were scaled on these dimensions. It was demonstrated that one dimension of classroom behavior--extent of controlling behavior--is predicted by scores on a questionnaire; the Teacher Preference Schedule, developed by Stern and Masling. Correlations with various behavioral measures of control ranged from .30 to .50 and were replicated in a second sample. The purpose of the present study is to investigate relationships between these dimensions of teacher characteristics and measures of pupil behavior.

This study, then, is designed to investigate relationships between teacher characteristics (both "behavioral" and "dynamic") and student responses. The behavioral characteristics of teachers include authoritarianism, warmth, and academic emphasis. The "dynamic" characteristics include achievement, affiliation, recognition, and control needs. The areas of student response which are of primary interest are those classified as academic achievement, creativity, group problem solving, and attitudes toward school. All treatises dealing with educational methods make assumptions as to relationships between teacher characteristics and student change but empirical evidence is virtually non-existent. Such evidence is imperative if one is to develop teacher training programs based on knowledge rather than opinion.

This report is divided, for convenience, into two parts: Phase I and Phase II. The procedures followed in each are presented in detail in following sections. In broad outline, however, the two phases were as follows.

Phase I

Sixty-five teachers for whom observational and test data were available from prior research and who were still teaching in the same schools were used. During the academic year 1961-1962 data was collected on the students of these teachers. The major categories of data were:

a. Achievement tests in reading and arithmetic at the beginning and end of the year.

b. Intelligence tests at the beginning of the year.

c. Selected "creativity" tests administered about two-thirds of the way through the school year.

d. Group problem-solving tasks administered about three-fourths of the way through the school year.

e. Orally administered questionnaire intended to measure such personality variables as dependence and achievement need as well as test anxiety and liking for school.

In addition the OSCAR device for evaluating teacher behavior was utilized for observing teachers approximately one-third of the way through the school year.

These data were treated as follows:

1. Pupil data were averaged within each class to provide a measure of class performance or gain which presumably reflects the influence of the teacher. Pearson correlations among these scores and indexes of teacher characteristics (both from our prior study and OSCAR) were computed both within grades and for all grades combined. The advantages of combining grades are (a) an increased number of cases and (b) it permits the analysis of the relationship between grade and the other variables. The disadvantages are (a) the possible obscuring of significant relationships which may differ from grade to grade and (b) the difficulty in interpreting relationships between variables which are themselves substantially related to grade. Partial correlation, however, is a partial answer to this problem.

2. Analysis of variance was employed in two ways. First, simple analyses were run comparing teachers (actually classes) within grades; (significant differences among teachers being a prerequisite to meaningful relationships with other variables. Second, specific hypotheses were tested through factorial designs using combinations of teacher and student characteristics as main effects and other pupil measures as dependent variables.

3. Although of secondary interest intercorrelations among the various pupil measures were analyzed by grade.

Phase II

A sub-sample of 19 teachers was selected for further pupil analysis. In grades two through five, four teachers were selected at the extremes of the controlling dimension, i.e., the two most and two least controlling in each grade were selected. In addition, two teachers in the second grade were selected as being high controlling but also high in affiliation or warmth though these characteristics are rather highly *negatively* correlated across the entire group of teachers. One additional high control teacher in the fifth grade was included as an alternate for certain analyses. During the academic year 1962-1963 the following data were collected on their students:

a. "Creativity" tests at the beginning and end of the year.

b. Intelligence tests at the beginning of the year.

c. Observations on a time sampling basis of four boys in each class paired on intelligence scores but the members of each pair differing greatly in total score on the creativity tests.

These data were treated as follows:

1. Differences in creativity gain under different teachers grouped as high and low on control were compared through analysis of variance.

2. In class behavior of high and low "creative" boys were compared in a design which also included the "controlling" teacher variable as a main effect.

3. Reliabilities of the "creativity" tests were determined with their correlations with the other measures.

CHAPTER II

REVIEW OF RELATED LITERATURE

Attempts to relate teacher characteristics or teaching methods to pupil measures have a long history. That most of the research has been quite unproductive attests to the complexity of the problem as well as to the lack of sophistication of the earlier approaches. In recent years what appear to be major strides have been made with respect to: (1) more objective measures of teacher behavior, (2) better measures of personality variables, (3) more refined techniques for dealing with measures of change or gain, (4) the development of new measures of pupil characteristics such as originality, group problem solving, and anxiety, (5) a more sophisticated approach to the problem which encompasses interactions among several variables as well as looking only for 'first order' relationships. As a result, a number of studies show promising findings.

The Measurement of Classroom Behavior

One of the first attempts to develop a behavioral measure of the social emotional climate in the classroom was made by Withall (1949). Withall developed a technique for recording and classifying the verbal behavior of the teacher in the classroom. The procedure was later modified and extended by Mitzel and Rabinowitz (1953), who visited classrooms in person and categorized teacher statements as they were given. The Withall technique employs a time sampling of teacher verbal behavior. A classroom observer records a verbal statement made by a teacher at the beginning of each one minute observational period. The teacher's verbal statements are categorized into six categories entitled Neutral, Reproving, Learner-Supportive, Climate Index, Problem Structuring, and Directive. The Climate Index provides a measure of the extent to which the classroom climate is learner-centered as opposed to teacher-centered.

A number of studies have been conducted to determine the reliability of the Withall technique as a measure of classroom behavior. Mitzel and Rabinowitz (1953), obtained reliabilities for the Withall categories as follows: Neutral .00, Reproving .00, Learner-Supportive .25, Climate Index .47, Problem Structuring .50, and Directive .50. These reliabilities were computed by means of an analysis of variance design based on assumptions which limited the generalizability of the findings to the particular sample of teachers studied and the particular observers used. A later investigation (Medley and Mitzel, 1958),obtained generally higher reliabilities for the Withall categories and was based on an analysis of variance design that permitted generalization to other teachers, observers, and situations. The reliabilities obtained in this later study were Problem-Structuring .79, Directive .51, Reproving .87, and Supportive .71. These investigators also obtained reliabilities of .81 and .79 for two categories of teacher non-verbal expressive behaviors entitled hostile and supportive respectively. The results of these studies also indicated that high observer agreement was obtained for two observers observing the same behavior (e.g., coefficients of observer agreement were frequently in the upper .90's). These results suggest that classroom observational procedures employing more than one observer in the classroom at the same time provide needless redundant information. If several observers are used they should visit the classrooms at different times to increase the sample of behavior observed.

Travers, et al (1961) employed a modification of the Withall technique to measure the classroom behavior of elementary school teachers. Teacher verbal statements were categorized into one of four categories entitled Academic Achievement Emphasis, Control, Affiliation, and Management. A sample of 77 elementary school teachers were observed by two observers working independently. Each observer collected two sets of 25 teacher statements. The reliabilities of the four categories of teacher behavior were estimated by means of analysis of variance. The reliabilities obtained were Academic Achievement Emphasis .50, Control .56, Affiliation .54, and Management .47.

These results indicate that the Withall technique provides a moderately reliable measure of the classroom behavior of teachers. Reliabilities of the magnitude reported are sufficient for research purposes. A modified version of the Withall technique was employed in the present investigation as a measure of teacher classroom behavior.

There have been a number of other attempts to develop reliable measures of classroom behavior. Cornell, Lindvall, and Saupe (1952), developed an observational system for measuring a number of dimensions of classroom behavior, such as the extent to which provision is made for individual differences, Social Organization, Initiative, Content, Variety, Competency, and Classroom Climate. Medley and Mitzel (1958b) conducted a reliability study of the Cornell dimensions and obtained reliability coefficients somewhat lower than those obtained for the Withall technique. The reliabilities obtained for the Cornell dimensions ranged from .00 to .42. The lower reliability of the Cornell procedure may have resulted from certain defects in the items on the observational schedule. A number of the items appear to require some subjective judgment on the part of the observer rather than the straightforward recording of observed behavior.

Medley and Mitzel (1958b, 1959) developed a measure of classroom behavior by combining and improving upon items from several earlier classroom observational techniques. The device, entitled the Observation Schedule and Record (OSCAR), provides measures of teacher behavior, pupil behavior, classroom grouping, educational materials employed in the classroom, subjects taught, and teacher verbal behavior. Twenty keys were derived from the measure, fourteen of which were reliable. The reliabilities of the fourteen keys ranged from .60 to .92. A centroid factor analysis of the fourteen variables derived from the OSCAR produced three factors entitled Emotional Climate, Verbal Emphasis, and Social Structure. The Emotional Climate dimension has been described by Medley and Mitzel as referring to the amount of observable hostility in the classroom as opposed to warmth and friendliness, Verbal Emphasis as the degree to which verbal activities predominate, and Social Organization as the amount of social grouping and pupil autonomy. The OSCAR device was employed in the present investigation as a means of measuring the classroom behavior of teachers and pupils.

The Prediction of Teacher Classroom Behavior

Ryans (1958, 1960) conducted an extensive series of investigations of teacher classroom behavior, and attempted to develop predictors of teacher behavior. The first stage of Ryans' investigations consisted of the development of a procedure for obtaining observer's ratings of teacher and pupil classroom behavior, and the factor analysis of these measures. The factor analysis of observer ratings of teacher and pupil classroom behavior produced five factors at the elementary school level and six factors at the secondary school level. Ryans derived three primary patterns of teacher classroom behavior from these factor analyses described as: (1) understanding, friendly, responsive vs. aloof, egocentrism; (2) responsible, businesslike, systematic vs. evading, unplanned, slipshod; and (3) stimulating, imaginative, original vs. dull, routine teacher behaviors. These dimensions of teacher classroom behavior are similar to the dimensions obtained from a factor analysis of teacher behavior measures conducted by Wallen, et al (1963). Seventy-seven elementary school teachers were Q-sorted on 24 dimensions of classroom behavior. A factor analysis of these Q-sorts resulted in five factors labeled as follows: cold, controlling vs. warm, permissive; vigorous, dynamic vs. dull, quiet; insecure, anxious vs. confident; spends much time alone vs. little time alone; and much academic activity vs. little academic emphasis.

Later stages of the Ryans studies attempted to predict the three criterion patterns of teacher classroom behavior from measures of teachers' attitudes, educational viewpoints, verbal ability, emotional adjustment, and validity of response (e.g., the desire of the teacher to make a good impression on the tests). The results indicated that teachers who exhibited much sympathetic, understanding classroom behavior expressed more favorable attitudes towards pupils and democratic classroom procedures. Sympathetic, understanding teacher classroom behavior was slightly positively correlated with the expression of more permissive, child-centered educational viewpoints; businesslike, systematic teacher behavior was positively correlated with traditional educational viewpoints; and stimulating teacher behavior was slightly positively correlated with liberal, permissive educational viewpoints. Low positive relationships were obtained between scores on a measure of verbal understanding and teacher classroom behaviors described as sympathetic, understanding, and stimulating. Finally, low positive correlations were obtained between a measure of emotional stability and sympathetic, understanding, and stimulating teacher behaviors. A low negative relationship was obtained between emotional stability scores and businesslike and systematic teacher classroom behavior.

Ryans (1958) attempted to investigate the behavioral correlates of teacher effectiveness as determined by principal's ratings of teacher performance. A group of "superior," teachers and a group of "poor," teachers were obtained by asking principals to select teachers from their schools who fell into these extreme categories. The comparisons of these groups on a number of dimensions indicated that "superior," teaching was most frequently associated with stimulating teacher behavior, businesslike teacher behavior, verbal ability, understanding friendly teacher behavior, and favorable teacher opinion of pupils.

The difficulties associated with the use of principals' ratings as a criterion of teacher effectiveness are well known. Wandt (1954) conducted a study of groups of teachers judged to be "superior" or "poor" teachers by their principals. One of the more interesting findings of this study was that teachers judged as "superior" by their principals scores consistently and significantly higher on a measure of attitudes towards administrators than did the teachers judged to be "poor" teachers. These data suggest to the reviewer that principals' ratings of teacher effectiveness may be contaminated by other variables associated with the teachers ability to get along with the principal, or the teachers acceptance of administrative authority in general. Many superior teachers, as defined by other criteria of teacher effectiveness, may endorse educational practices or viewpoints which conflict with the more traditional viewpoints of school administrators. Better criteria of teacher effectiveness would be provided by more objective measures of changes in pupil behavior, or gains in pupil achievement.

Several studies have attempted to predict the classroom behavior of elementary school teachers from measures of the teachers' needs. Wallen, et al (1963), obtained a number of criterion measures of teacher classroom behavior including measures of observed teacher classroom behavior, observers' ratings, and observors' Q-sorts of teacher behavior. The predictor measures in these investigations were a paper and pencil projective device designed to assess the need imagery projected by a teacher in response to the behavior of some hypothetical person, and the Stern and Masling Teacher Preference Schedule measure of needs (Stern and Masling, 1958). The latter device is a straightforward questionnaire consisting of 100 statements to which the respondent indicates the extent of her agreement. Both instruments were keyed to provide four scores for achievement, affiliation, control, and recognition needs. Although the projective measure of needs generally failed to predict teacher classroom behavior in these investigations, the Stern and Masling Teacher Preference Schedule measure of control need consistently predicted teacher classroom controlling behavior in two samples of elementary school teachers.

Wodtke, et al (1963) in an extension of the preceding study attempted to predict teacher classroom behavior using a measure of the pattern of teacher needs derived from the Stern and Masling Teacher Preference Schedule. The study attempted to test the hypothesis that teacher classroom behavior related to a particular need depends, not only on the absolute strength of the need but also on the relative strength of that need with respect to other needs. According to this hypothesis, a teacher with a high control need might show little behavior related to this need if the teacher were characterized by an even stronger need for affiliation. The results indicated that an affiliation need pattern score predicted teacher classroom behavior somewhat better than the conventional need score. A similar analysis of the control need pattern score indicated somewhat poorer prediction from the pattern score. These findings suggest that the relative strength of needs may determine the behavior emitted to satisfy some needs, but that other needs may function independently.

The Effects of Teacher Characteristics and Teacher Classroom Behavior On Pupil Behavior and Achievement

One objective of studies of teacher characteristics and teacher classroom behavior is to determine the effects of such behavior on criteria of pupil behavior and achievement. Prior to 1950 there was little agreement as to what constituted effective teaching. During the past ten years there has been a growing emphasis on behavioral measures of criteria of teacher effectiveness such as measures of change, gain, or growth in student achievement.

Mitzel (1960) reviewed studies of the criterion problem of teaching effectiveness and categorized the criteria which have been used into three main classes: product criteria, process criteria, and presage criteria. Product criteria may be defined as changes in the behavior of pupils. An example of a product criterion of teaching effectiveness is measured pupil achievement gain. The advantages of product criteria of teacher effectiveness have been discussed by Rabinowitz and Travers (1953), and Remmers (1952, 1953). Generally, the advantages of such criterion measures are their greater reliability, objectivity, and reproducability, and their freedom from bias commonly associated with such measures as principals' ratings of teacher effectiveness. The ultimate criterion may be the ability of the student to function as an adult member of society. It is assumed that students who exhibit superior achievement in subjects such as reading, arithmetic, social studies, etc. will be better able to function in that capacity.

The other classes of teacher effectiveness criteria discussed by Mitzel are several steps further removed from the ultimate criteria and require additional assumptions concerning the intervening steps. For example, process criteria refer to aspects of teacher and pupil behavior which are believed to be worthwhile in their own right, such as teachers' disciplinary practices, the ability of the teacher to maintain rapport with students, and the degree to which teachers individualize instruction. The use of such criteria requires further test of the assumption that good teacher-pupil rapport does in fact produce pupils who learn more subject matter, who are better problem solvers, who are better adjusted socially, etc. Presage criteria refer to criteria based on precedent and "common sense." Thus, it is assumed that a well adjusted teacher of good character and high intelligence is a more effective teacher. Since criteria of this type must still be checked for their effects on pupil behavior, the present reviewers would argue that a more efficient procedure would be to go directly to criterion measures of pupil change.

The measures employed in the present investigation were outlined in the introductory statement. Generally, the criterion measures were of the product cirterion variety such as pupil achievement gains, measures of pupil problem solving, and pupil anxiety. In several instances process criterion measures were employed such as measures of pupil classroom behavior.

Significant differences among teachers have been found for criterion measures of pupil achievement gain. Mitzel and Medley (1957) examined the reading achievement gains of the pupils of 49 elementary school teachers in the New York City public schools. These investigators introduced a statistical procedure for the study of pupil gains which appears to have much promise for studies of teacher effectiveness. Equivalent forms of the California Reading Test were administered in the fall and in the spring. A pupil's reading gain score was the difference between his end of year score and the score predicted for him on the basis of his performance at the beginning of the year. This procedure provides a control for differences in previous reading achievements among pupils. Significant differences were obtained among teachers in the reading gains achieved by their pupils which could not be attributed wholly to differences among schools or to pupils' ability or previous achievement.

Medley and Mitzel (1959) attempted to relate five measures of teacher effecitveness, including a measure of pupil achievement gains, to three dimensions of teacher classroom behavior. The five criterion measures of teacher effectiveness were average adjusted pupil reading growth, pupil growth in group problem solving skill as measured by the Russell Sage Social Relations Test, pupil-teacher rapport as measured by a short pupil questionnaire, supervisors' ratings, and teachers' self-ratings. The three dimensions of classroom behavior entitled Emotional Climate, Verbal Emphasis, and Social Organization were derived from the OSCAR observational technique which was described above. The measured pupil gains in reading achievement and in group problem solving skill did not relate to the measures of teacher and pupil classroom behavior. The measure of teacher-pupil rapport was found to relate to the behavioral measure of emotional climate. Supervisors' ratings correlated positively with the friendliness of the teachers' classrooms, but did not correlate with pupil growth in reading or problem solving skill. These latter findings emphasize some of the difficulties associated with supervisors' ratings as a measure of teacher effectiveness.

Silberman (1957) attempted to determine the effects of differences among teachers in their frequency of using praise and reproof on pupil gains in reading achievement. A measure of teachers' praise and reproof was obtained by means of the Withall observational technique. The frequency of praise used by the teacher was the number of supportive statements emitted by the teacher in the classroom. The score for reproof was the number of reproving teacher statements which were recorded by classroom observers. No significant relationships were obtained between the measures of teacher praise and reproof and a measure of pupil gain in reading achievement which was adjusted for the pupils' ability and previous reading achievement. The lack of significant relationships in this study suggests that the effects of teacher praise and reproof on pupil learning may depend upon other differences among teachers and pupils. For example, interactions may exist between teacher praise and reproof and pupil personality variables such as anxiety, achievement need, affiliation need, etc. which determine the extent of pupil learning. One objective of the present investigation was to determine the effects of interactions between teacher classroom behavior and pupil personality variables on measures of pupil achievement gain.

Christensen (1960) studied relationships between teacher warmth and permissiveness, pupil affect-need, and pupil growth in achievement. The measures of teacher warmth and permissiveness were obtained from a questionnaire concerning various classroom practices administered to pupils in the fourth and fifth grades. The measures employed in this study might best be described as measures of teacher warmth and permissiveness as perceived by pupils. The measures of teacher permissiveness and pupil affect-need did not relate significantly to pupil achievement gain, nor were any significant interactions obtained. Perceived teacher warmth, however, was significantly related to pupil gains in vocabulary and arithmetic achievement.

The above study raises a question concerning possible differences between measures of teacher classroom behavior as perceived by pupils, and measures of teacher classroom behavior as perceived or recorded by outside observers. The present investigators, having spent extended periods observing teachers in classrooms, are convinced that measures of teacher behavior as perceived by students may provide results different from those obtained from observers measures of teachers classroom behavior. The present investigation provides data on this point since measures of teacher classroom behavior as perceived by pupils were obtained in addition to measures obtained by classroom observers. Correlations between the pupil measures and observer measures will indicate the extent to which these measures provide similar results.

Cogan (1958a, 1958b) conducted an investigation of relationships between teacher classroom behavior as perceived by pupils, and pupil required and self-initiated work. The sample consisted of 33 teachers and 987 eighth grade pupils. Three dimensions of teacher classroom behavior were measured by means of a questionnaire administered to pupils. Inclusive teacher behavior consisted of the extent to which the teacher served as a cue to approach behavior tendencies on the part of pupils such as "liking," or "respect." The inclusive behavior dimension included integrative, affiliative, or nurturant teacher behaviors. The preclusive dimension of teacher behavior indicated the extent to which the teacher was perceived by pupils as a cue for anxiety or avoidance, and included dominating, aggressive, and rejecting teacher behaviors. The conjunctive dimension indicated the level of the teacher's demands on his pupils, his ability to communicate with his pupils, and his competence in classroom management. The two dependent variables in the study were the pupils' estimate of the amount of work required of him, and the extent to which he engaged in self-initiated work. The data were analyzed in two ways. The "perception" analysis examined within class correlations between the three dimensions of teacher behavior and the two measures of pupil work output. The "trait" analysis examined averages of the measures among the classes.

The results of the "perception" analysis failed to confirm the hypothesis that preclusive teacher behavior as perceived by pupils would be negatively related to the amount of required and self-initiated pupil work. The analyses of the inclusive and conjunctive teacher behavior dimensions, however, confirmed the hypotheses that these dimensions would be positively related to the amount of required and self-initiated pupil work. Some interesting additional findings were that principals' ratings of teacher behavior generally did not relate significantly to the pupils' perceptions of teacher behavior, but the teachers' estimates of the pupils' required and self-initiated work were found to correlate positively and significantly with the pupils' own estimates on these variables.

The results of Cogan's "trait" analysis (e.g., the among groups correlations between the average scores on the three teacher behavior dimensions and the average scores on the measures of required and self-initiated pupil work), indicated that teacher inclusiveness as perceived by the pupils was positively and significantly correlated with the pupils' estimates of their required and self-initiated work output. Cogan's study supports the hypothesis that teachers who are seen by pupils as more approachable and likeable require more work of their pupils, and generate more self-initiated work on the part of their pupils.

One final study of the effects of teacher behavior and personality on pupil achievement gains remains to be mentioned. Heil, et al (1960) identified three main teacher types, and four types of pupils on the basis of personality tests. The sample employed in the investigation consisted of 55 elementary school teachers in grades four through six, ane their pupils in the Brooklyn, New York, public schools. The three teacher personality types were the turbulent teacher, the self-controlling teacher, and the fearful teacher.

The turbulent teacher was described as one whose aggressive and sexual impulses appear to be close to the surface, who had a low need for approval and acceptance by others, and who showed a high preference for thinking in the mystery, life-death-universe and fantasy categories. The turbulent teacher was described as unmethodical and verbally aggressive in social interactions. Such a person would most likely irritate his peers, and experience frequent clashes with authority figures. The turbulent teacher had a high preference to be independent of direction by authority figures and a need to do uninhibited and unconventional thinking.

In contrast to the turbulent teacher, the self-controlling teacher was described as feeling most secure when things were running smoothly, and evidenced a high preference for being methodical. The self-controlling teacher appeared to spend a considerable amount of energy keeping his impulses under control, and was more concerned with what other people thought of him, and less introspective. The self-controlling teacher also appeared to be submissive in relation to authority figures, but would tend to act in an authoritarian manner to his subordinates. He would be more likely to react to others with terms such as "good" and "bad" and would probably be better liked b his ppers than the turbulent teacher.

The third teacher type, or fearful person, was described by Heil as one who fears many aspects of the environment. Such a person would evidence a fear of sexual impusles, a fear of being alone, a fear of mystery, and feelings of dependence and helplessness. A fearful teacher would tend to reveal rigidity of thinking in the face of uncertainty, and a tendency to be consistently protecting himself in his relations with others.

The four pupil types identified were conformers, opposers, waverers, and strivers. The conforming pupils were characterized by incorporation of and conformity to adult standards, a high social orientation, strict control of impulses, and an emphasis on mature behavior. The outstanding characteristics of the opposers were disturbed authority relationships, oppositional tendencies, pervasive pessimistic tone, intolerance of ambiguity, and a disappointment in self and others. The waverers were characterized by a high level of anxiety, ambivalence, fearfulness, and indecisiveness. Finally, the strivers were characterized by a strong attitude of striving, a drive for recognition with regard to school achievement, and exhibitionistic needs.

Other measures obtained in the study were observers' ratings of the teachers' classroom behavior, a measure of the teachers' professional knowledge, pupil achievement gains over one school year adjusted for pupil intelligence, and pupil gains on a measure of friendliness towards others.

The results of the study indicated that the self-controlling teacher type generally produced greater gains in pupil achievement than did the turbulent and fearful teachers. The effectiveness of the self-controlling teachers relative to the turbulent and fearful teachers was particularly marked for opposing and wavering pupil types. The turbulent and fearful teachers were generally more effective with conforming children than with opposing and wavering children. When the achievement gain was compared for the four pupil types, the strivers evidenced the greatest gain, and did so regardless of the teacher's personality type. The lowest mean achievement gain was attained by the waverers.

When pupil gains in friendliness were compared for the three types of teachers, the self-controlling teachers again produced significantly greater gains than did the turbulent and fearful teachers. A comparison of the friendliness gains of the four pupil groups produced non-significant results, although there was a tendency for the strivers to achieve the highest gains.

Significant between schools effects were obtained for the above dependent variables, however, the results attributed to teacher and pupil differences held up when the dependent variables were adjusted for such effects.

Several additional findings were of interest. The observers' ratings of teacher classroom behavior failed to correlate with mean class achievement gains. In addition, the teachers' scores on the measure of professional knowledge failed to correlate significantly with class gains in achievement.

An important contribution of Heil's study, in addition to the main findings, is it's emphasis on the measurement of both teacher and pupil variables. Although the main focus of the study was not on possible interactions between teacher personality variables and pupil personality variables as determiners of pupil achievement, such interactions can be seen by inspection in several tables of results. The fact that the three teacher types were not equally as effective or ineffective across the four pupil types indicates the presence of significant interaction effects. Placing primary emphasis on teacher variables or pupil variables alone and failure to examine the interactions of such variables may be a primary cause of the inconsistency of results in studies of teacher effectiveness. There is little reason to expect that one teacher type, or one class of teacher classroom behavior will be effective with all types of pupils. Recent studies of the effects of pupil anxiety on pupil achievement (Phillips, 1962; Cox, 1960) point to the importance of pupil anxiety as a classroom variable. One of the major objectives of the present investigation was to set up an experimental design which would enable the investigators to test for possible interactions between the level of teacher classroom control and such pupil variables as anxiety, dependency, need affiliation, need achievement, etc.

CHAPTER III

PHASE I--PROCEDURE

Sample of Teachers

Of the original group of 118 female teachers utilized in the prior study of teacher behavior 65 participated in the present study. The decision not to include kindergarten or sixth grade reduced the number by 18. The remaining 35 were no longer teaching in the same schools. In some analyses the N is reduced slightly due to the following:

1. One teacher requested that she not be observed further during class.
2. One teacher left school mid-way through the year.
3. One teacher was ill during part of the year.
4. Due to specialization within certain schools some fourth and fifth grade classes received instructions in reading and/or arithmetic from other than their major teacher. The number of classes so treated was two for reading and five for arithmetic.

The sixty-five teachers represented six schools; the N per school ranged from six to sixteen. They varied considerably as to age and experience. The six schools were selected so as to provide a variety of socio-economic background.

Teacher Measures

The following scores were utilized as measures of teacher characteristics:

1. Obtained during previous study (years 1959-60 and 1960-61). For some teachers these scores were obtained two years prior to the present study; for others one year prior to the present study.
    a. Scores intended to reflect four needs as follows: control, affiliation, achievement, and recognition. Four paper-and-pencil tests were used:
        (1) A questionnaire regarding preferences for various activities related to teaching and to children (TPS).
        (2) A projective test requiring explanations of examples of behavior (USB).

(3) A questionnaire regarding attitudes about educational objectives (Obj.).

(4) A series of classroom situations to which the respondent evaluated various teacher reactions (Sit.).

Thus a total of sixteen scores were obtained--four needs across four tests.

b. Scores intended to reflect teacher behavior in the classroom. Our primary interest was in the following postulated dimensions of teacher behavior:

(1) Achievement: degree of emphasis on academic activities and the importance of excellence.

(2) Affiliation: degree of interaction with students and emphasis on warm personal relationships.

(3) Control: degree to which the moment to moment behavior of students is controlled by the teacher.

(4) Recognition: degree to which the teacher is the center of attention.

Two means of assessing these dimensions were used. The first is based on sampling of the verbal behavior of the teacher; the second consists of observer ratings and elaborations thereof. Each of the participating teachers was visited in class by two members of the research team. After preliminary visits to reduce the effect of an observer, data collection was begun. Each teacher was visited at least three times by each observer.

(1) Verbal Statements. This technique requires the verbatim recording of teacher statements based on time sampling. The statements were subsequently categorized and categories combined so as to provide a frequency score of statements in four categories: controlling, affiliative, achievement, and management. Too few 'recognition' statements were recorded to warrant using this score.

(2) Ratings. Each teacher was rated by each of two observers on a total of 35 dimensions gleaned from prior research. In the present study only ratings on the four dimensions of major interest (control, affiliation, achievement, and recognition) were used.

(3) Q-sorts. Each teacher received a Q-sort score based on a combined sort of two raters working together. Twenty-four characteristics, selected from the 35 used in the ratings, were sorted. A factor analysis of these sorts resulted in five factors as follows:

a. Cold, controlling vs. warm, permissive

b. Vigorous, dynamic vs. dull, quiet

c. Insecure, anxious vs. confident

d. Works alone vs. interactive

e. Much academic emphasis vs. little academic emphasis

Each teacher was assigned a score on each factor which was the sum of the Q-sort scores on the dimensions having highest loadings on each factor.

A detailed description of the need measures and the behavior measures may be found in our earlier report (Travers, Wallen, Reid, and Wodtke, 1961).

2. Teacher Measures Obtained During the Present Study (1961-62).

a. OSCAR. This device, developed by Medley and Mitzel, has shown promise as a means of quantifying differences in teacher behavior. During each observation period the observer samples time periods of five minutes duration, alternating his focus of attention until a total of six periods (30 minutes) are completed. During half the period the observer records whether or not certain events occur, i.e., children are in small groups. During the other periods a frequency count is made of teacher statements classified in categories. Fourteen dimensions plus three factors are scored. The scores obtained were as follows. Certain specific categories are also analyzed.

Fourteen variables:

(1) Time spent on reading

(2) Frequency of problem structuring teacher statements

(3) Autonomous administrative groupings--groups set up by teacher for academic work

(4) Pupil leadership activities (frequency)

(5) Freedom of movement

(6) Manifest teacher hostility

(7) Supportive teacher behavior

(8) Time spent on social studies

(9) Disorderly pupil behavior (frequency)

(10) Verbal activities

(11) Traditional pupil activities (board work, recitation, reports)

(12) Teachers verbal output

(13) Audio-visual materials

(14) Autonomous social groupings

Factors:

(1) Emotional climate (teacher hostility, disorderly behavior)

(2) Verbal emphasis (time spent on reading, verbal activities)

(3) Social structure (autonomous social groupings, teacher verbal output)

Each teacher was visited once by each of three observers. Two visits per observer were originally planned but time did not permit this much observation.

b. Q-sort. Inasmuch as a time period had elapsed since the previous Q-sort scores were obtained, and as a check on the original scores, each observer independently sorted the present sample of teachers on one of the variables which had been most promising in the prior study, i.e., affiliative behavior. The resulting correlation was .57. Considering the fact that the observers making the second sort included one who had no prior contact with the teachers and two who had prior contact with approximately one-third of the teachers and that the observer time prior to sorting was limited to one observation of the teacher, this correlation indicates considerable stability of this dimension of teacher behavior. Subsequent selection of high and low controlling teachers excluded those for whom disagreement between sorts was high. For this selected sub-sample the correlation between the two sorts was .87.

c. Pupil Measures. Immediately following the administration to the pupils of the Russell Sage Social Relations Test (discussed subsequently) the teacher's reaction was assessed after discussion on the part of the two observers. Each teacher was assigned a rating as follows:

(1) Extent of ego-involvement--4 point scale
1 = left room
2 = uninvolved
3 = mildly involved
4 = highly involved

(2) Extent of ego-involvement--2 point scale
1 = left room or uninvolved
2 = involved

(3) Nature of reaction

1 = strong disapproval
2 = mild
3 = neutral reaction
4 = mild positive reaction
5 = strong positive reaction

(4) Degree of defensiveness re class behavior

1 = none
2 = mild
3 = severe

Pupil Measures (1961-62)

1. Intelligence. California Test of Mental Maturity--Short Form--Primary Level--in grades 1-3; Elementary Level in grades 4-5. Administered in the fall. Scores used: Language, Non-language, and Total.

2. Achievement. California Achievement Tests: Reading Vocabulary, Reading Comprehension, and Arithmetic Fundamentals. Administered fall and spring as follows:

	Fall	Spring
Grade I	Lower Primary Level	Lower Primary Level
Grade II	Lower Primary Level	Lower Primary Level
Grade III	Upper Primary Level	Elementary Level
Grade IV	Elementary Level	Elementary Level
Grade V	Elementary Level	Elementary Level

3. "Creativity." The following tests developed by Torrance were used in the grades indicated. The scores derived from each are listed.

a. Incomplete Figures Test (Grades 1-5)

(1) Originality
(2) Closure
(3) Productivity

b. Circles Test (Grades 1-5)

(1) Relevance
(2) Redundance
(3) Ideational Fluency
(4) Flexibility
(5) Originality

c. Consequences Test (Grades 3-5)

(1) Fluency
(2) Flexibility
(3) Originality

4. Questionnaire. (Read orally--S responded yes or no to each item.) (Grades 1-5) With the exception of (a) and (h) these scales were constructed as part of the present project.

a. Sarason Test Anxiety Scale (30 items)

b. Dependency Need (10 items)

c. Affiliation Need (10 items)

d. Achievement Need (10 items)

e. Perception of teacher as warm, affiliative (5 items)

f. Perception of teacher as controlling (5 items)

g. Perception of teacher as emphasizing academic matters (5 items)

h. Liking for school (5 items) (Taken from 'My Class Inventory,' Medley and Klein, 1957)

Each of the preceding devices was administered to all children (less absentees) in the grades indicated. In addition, the Russell Sage Social Relations Test was administered in grades 2-5. This test requires the group to devise and execute a plan for constructing of figures out of blocks to match paper templates. A more detailed description taken from the author's materials follows:

> 5. "The Russell Sage Social Relations Test is a situational test created to assess the nature and quality of two important aspects of elementary school children's skill in social relations: (1) skill in cooperative group planning procedures, and (2) skill in techniques of cooperative group action. It is designed for children in grades three through six and occupies approximately one hour of time. Its administration requires a trained examiner and a trained observer.
>
> The test consists of three construction-type problems, graded in difficulty and administered one after the other. For each problem the children are provided with thirty-six interlocking construction blocks of various shapes and colors and a model which they are to copy exactly. All thirty-six blocks are necessary for the construction of each model, there being none left over when the problem is completed. The first problem requires the children to build the figure of a house; the second, a simple footbridge; and the third, the figure of a dog.
>
> The test is administered to an entire classroom group of children at once. In the course of the administration, the examiner tells the children that the purpose of the test is to see how well they can work together. He tells them that the test does not provide scores for individual pupils, but only one score for the class as a

whole. Pupils may, therefore, help each other, discuss things freely, and work cooperatively on the problem. He provides each child in the room with one or two of the blocks needed to build the particular model shown them. He then tells the group that, although they can have all the time they need to plan how they are going to go about it, they will be allowed only fifteen minutes for the actual construction. Finally, the examiner informs the group that their score on the test depends upon the time it takes them to build the model, the shorter the time the higher their score.

Having explained the test and its rules to the class, the examiner allows the pupils to proceed with the task of developing a plan of action. When this plan is completed and all members of the group have it clearly in mind, he tells them to begin construction of the first model and immediately starts timing. He stops his watch when the final block is correctly placed and announces the score to the class. He then proceeds with the second and more difficult problem. The class is again given time to plan and organize itself as in the first problem. The third and most difficult problem is administered in like manner.

As mentioned above, the test is designed to provide information about two different aspects of social relations skill. The first is the ability to participate in group discussions of plans for future group action; the second is the ability to regulate one's behavior in accordance with these plans so that movement toward the agreed-upon group goal is facilitated. The test is thus divided into two separate parts. The first part is termed the Planning Stage--the period during which the children devise a plan of action. The second part is called the Operations Stage--the period during which the group puts its plan into effect and actually builds the model, using the blocks provided by the examiner.

In both stages the aim is to obtain a measure of the best possible performance the group is capable of rendering. This aim is accomplished by having the examiner play a precisely defined and standardized role in his administration. His initial task is to explain the test and its rules to the children. In fulfilling this task his actions and statements are the same for all groups. His second and more complicated assignment is to create and maintain an atmosphere in which optimum group planning is possible. To achieve this purpose he must vary his behavior in accordance with the needs of the group. That is, a noisy group must be held in check so that communication and discussion of ideas is possible; a quiet solemn group must be encouraged to express ideas; a group that gets lost in the complexity of its own ideas must be rescued and its thinking clarified. Throughout all this the examiner must refrain from providing the children with his own ideas of ways of solving the problem and from telling them which of the ideas they have suggested are good or poor. He must at all times maintain a permissive atmosphere in which the giving of ideas is encouraged and all suggestions warmly accepted, but in which no positive direction is supplied. In such an atmosphere the children can perform only at the level at which they themselves are presently capable, and the quality of their planning thus provides a measure of the amount of skill which they possess.

The examiner's final task is to insure that, before entering the Operations Stage, each child in the group clearly understands the plan which has been devised. Therefore, he repeats verbatim the details of the plan which the class has built and asks if there are further questions before they begin the construction task. Here again the examiner must vary his behavior with the needs of the group, answering questions on points which the group has already decided, and referring back to the group any questions about points which they have not taken into account. It is not until the children tell the examiner that they are all ready to begin work that he permits them to start. This insures that all groups enter the Operations Stage with presumably equal clarity concerning the plan that they as a group have built.

During the Operations Stage the role of the examiner is also standardized. After the signal to start is given he withdraws to one side of the classroom and refuses to answer questions or provide help, even when requested. Regardless of what the children do (however noisy, rough, boisterous, or quarrelsome they become, or however quietly and efficiently they work), he neither reprimands nor commends them. He interferes only if behavior gets so out-of-hand that physical harm to children is apt to result. If this occurs, he stops the test and no further problems are given.

The mechanics of administering the test place the children in a situation where progress toward the goal is not possible without some type of cooperation on the part of each child in the group and where external controls designed to elicit particular kinds of behaviors and suppress other kinds do not exist. In such a situation the children can show the extent to which they are capable of carrying out a plan of action they themselves have devised, and the kind of behavior they exhibit provides a measure of the amount of skill they possess in this respect.

During the time the children are engaged in the task of planning how to solve the problems and in working on the actual construction of the various models, the observer (seated near the back of the classroom) keeps a record of their behavior using standardized observation sheets. The data thus collected are later transformed into a set of numerical scores which are used to rank groups on the basis of their skill in cooperative group planning procedures and in techniques of cooperative group action."[1]

Scoring. The obtained protocols are scored on point scales for the following dimensions:

a. Planning Stage.

(1) Extent and quality of participation

---

[1] Damrin, D. E., Journal of Experimental Education, 1959, p. 85.

(2) Quality of communication

(3) Quality of ideas

(4) Quality and precision of final plan

(5) Extent of involvement (as distinguished from active participation

(6) Extent of group autonomy or independence from examiner

b. Operations Stage.

(1) Extent of involvement

(2) Atmosphere (extent of hostility vs. cooperation)

(3) Activities of non-involved members

(4) Success or task accomplishment

The test as used in the present project departed from standardized procedure in two ways. First, time permitted using only two of the three problems. The second, and more serious departure arose from lack of clarity in the manual as to whether subjects are permitted to use more blocks if they so desire. Although the article quoted above implies they are not, we had previously made the contrary decision and allowed additional blocks to be used. It is likely that this affected the task in the following ways:

1. It fostered scrambling for blocks and hence greater disorder in some classes;
2. It permitted the children to set up competing groups (e.g., boys vs. girls) each of which worked on the design.

In general, our procedure seems less conducive to efficient, democratic procedures, making some of the group performances all the more striking in their adequacy of solution.

Administration and Scoring of Measuring Instruments

The intelligence and achievement tests in the fall were administered during a four week period during the month of October. All testing was done by the research staff or graduate students in the Department of Educational Psychology. In grades 1 and 2, two testing periods per class were utilized, during one of which the California Test of Mental Maturity was administered; during the other of which the achievement tests were given. In grades 3, 4, and 5 three sessions were utilized, one for the CMM, a second for the reading achievement test and a third for the arithmetic achievement.

The achievement testing in the spring of the year was conducted in the same fashion during a period from approximately April 15 thru May 15. The questionnaire measure including the Test Anxiety Scale was administered during the latter part of January during approximately a two-week period after previous try-out at the campus laboratory school. The administration of this device took approximately half-an-hour, during which time each question was read aloud and the children responded with a yes or no answer--IBM answer sheets were used in grades 3 thru 5 and a form enabling the child to circle yes or no was used in grades 1 and 2. In the first two grades the classes were divided so as to work with no more than fifteen children in one group.

The in-class observations utilizing the OSCAR device were accomplished during the period from February 1 thru March 15 with the exception of a couple of observation periods which had to be scheduled later. Each of the three staff members visited each teacher once. As previously mentioned it was anticipated that two or more visits per observer would be made but time did not permit this much observation.

The creativity tests were administered during the latter part of March during approximately a three-week period from March 15 on. Once again the lower two grades were sub-divided to provide smaller groups for this testing. The Russell Sage Test was administered during the period from April 1 thru May. The principle investigator was the test administrator in all classes and one of the research assistants fulfilled the role of observer in all classes.

The scoring of all objective tests was accomplished by utilization of test scoring equipment for the upper three grades wherein IBM answer sheets were appropriate or by hand-scoring the materials in the first two grades. This applies to the CMM, CAT, and questionnaire data. The creativity tests were scored by two individuals both experienced in psychological test scoring under the supervision of one of the research assistants. Torrance's Manual was followed with several clarifications obtained through correspondence with Torrance. Random samples for each test varying in size from N = 86 to N = 121 were used to determine the agreement between the scorers. The agreement was as follows:

Circles Fluency	.84	Incomplete Figures Originality	.92
Circles Redundancy	.87	Incomplete Figures Closure	.87
Circles Originality	.87	Consequences Flexibility	.86
Circles Flexibility	.96	Consequences Fluency	.95
		Consequences Originality	.81

Scoring of the Incomplete Figures-Productivity is completely objective. In spite of this rather high agreement between scorers, there were consistent differences in means ascribed by the scorers on several of the scores for the Circles Test and hence an adjustment was necessary to compensate for this as discussed in the section on data analyses. The scoring of the OSCAR device utilized the procedures recommended by Medley and Mitzel. This procedure is completely objective involving either the absolute number or percentage of responses checked in various categories of the device.

Scoring of the Russell Sage device also followed the procedures recommended in the manual with the addition of scores added by the present researchers. The scores which were added were scores intended to reflect, in the planning stage, extent of participation, adequacy of communication, degree of autonomy of the group, and, during the operations stage, scores intended to reflect the variation in atmosphere under the categories (1) quiet vs. noise, (2) friendly vs. hostile. The basis for this scoring is found in Appendix A. These protocols were scored by two members of the research staff for the total number of classes wherein this device was administered which was .50. The scoring agreement expressed as a correlation is as follows:

Planning Stage:

Participation	.85
Communication	.96
Ideas	.91
Plan	.80
Involvement	.94
Autonomy	.56

Operations Stage:

Involvement	.93
Atmosphere	.85
Quiet-Noisy	.83
Friendly-Hostile	.78
Activity	.42
Success	.93

These correlations indicate the degree of agreement between scorers on problem number one only. It was also necessary to check on the agreement obtained by the two observers during the operations stage. This was obtained by using as the score the sum of the score accorded by the two scorers for each observer protocol. The correlations between the two observers then were as follows:

Operations Stage:

Involvement	.79
Atmosphere	.80
Quiet-Noisy	.85
Friently-Hostile	.79
Activity	.59

It was not possible to check on observer evaluations during the planning stage since only one observer was recording this information. It appears, however, that the recording during the planning stage is considerably more objective than it is during the operations stage.

Data Analysis

The first analysis consisted of obtaining frequency distributions, means, standard deviations, and intercorrelations for each of the student measures within each grade. This analysis served two purposes. First, it was of interest to examine these data, especially the intercorrelations, in their own right. Second, it was necessary to examine the distributions of several measures--achievement and intelligence--in order to decide how best to obtain scores having the same meaning across the various levels of the tests used. Further, certain correlations were necessary in order to obtain several of the major scores of interest, i.e., the difference between spring score as predicted by a combination of intelligence score and fall score and actual spring score in the various achievement areas. The latter point requires additional discussion.

One of the principal measures of interest in this study is the score reflecting gain in the academic content areas of reading vocabulary, reading comprehension, and fundamentals of arithmetic. There are numerous reasons why a simple gain score is less desirable than a difference score taken as a difference between end-of-year predicted score and actual score. Thus, in each of the three content areas, multiple regression equations were established within each grade using as predictor variables, CMM total score and the pre-test or fall score on the achievement test. The predicted score was the end-of-year or spring score.* After these regression equations were obtained they were used to predict the end-of-year score for each student in each of the content areas. This score was then subtracted from the actual score giving us the measure of gain. This score was then punched into the IBM card which contained the other scores for each student.

Following this the data decks were submitted for correlational analysis within each grade, using an incomplete data program. In addition, means and variances were obtained from these data both for the grade as a whole and for separate classes or teachers within each grade.

The next analysis to be carried out was an analysis of the teacher matrix as opposed to the previous analysis based on scores for individual pupils. In the teacher matrix all of the prior information which we had obtained on a given teacher was included as well as the average of the pupil measures for each class, i.e., teacher. Thus, the mean gain on the achievement tests, the mean creativity score, the mean score on the anxiety scale, etc. are entered as scored for each class and are viewed as scores pertaining to the particular teacher. All teachers (across the five grades) were used in this analysis. On the creativity measures and the questionnaire measures it is meaningful to simply analyze the data in terms of raw scores (it might be argued that some form of scaling would improve the meaningfulness of these scores but this was not attempted at this time). In the case of the achievement measures and the CMM, however, simply using the raw scores or one of the manual conversions from the raw score (such as IQ or Grade Equivalent) seems indefensible due to the fact that different levels of the test were utilized across the five grades and there is no assurance that a given score has the same meaning at one grade as it does in another. In fact, it is quite obvious in many cases that it does not. This, of course, causes no problem for the pupil matrix analysis or within grade teacher matrixes which are done within each grade. However, for the composite teacher matrix it is necessary for the mean value used to represent the performance of the class to have the same meaning across the various grades. The simplest way to handle this problem is simply to convert the mean value for each teacher to a standard score based on the entire distribution of scores for that grade. The only possible bias introduced by this procedure occurs if the distributions within grades are quite different. Too, if the distributions were about equal in variability from grade to grade, then one could simply use a standard score for each teacher mean and it would have comparable meaning from grade to grade. If, however, the distributions showed differences in variability from grade to grade, then a standard score would not have comparable meaning across grades. Although the distributions for the gain scores were all unimodal, pretty symmetrical and not too discrepant from grade to grade there was enough difference to warrant normalizing. Consequently, the score that we used for each of the teacher mean gain scores was a normalized standard score based on the total distribution of scores. We took the percentage of students within each grade with scores below the particular mean score and used this percentage value to obtain the corresponding standard score. This z score was then transformed to a scale using a mean of 50 and a standard deviation of 10. In addition to this score, which is one of the scores used in the analysis, we also used a second score for each teacher for each of the gains which is an attempt to partial out the effects of time lag since we discovered that the time lag between testings was correlated low but significantly with gain score. The exceptions to this were the first grade reading vocabulary, the fourth grade reading vocabulary, and the fifth grade arithmetic wherein the correlations between gain and time lag were insignificant. The procedure followed in this adjustment was as follows: Within a grade the time lag scores were converted to standard scores with a mean of 50 and a standard deviation of 10 but not normalized. Using the within grade correlations, gain score was predicted on the basis of its correlation with time lag. It was assumed that the correlation would be the same if each of the gain scores had been normalized, although this was not done since it would have required converting each score to a standard score and then computing the correlations--a laborious procedure. Then the predicted gain score in standard score form was subtracted from 50, (the mean of the predicted scores) to obtain a discrepancy of the predicted from the mean. This quantity was then subtracted from the original obtained gain score (again in normalized standard score form) when the predicted scores were above the mean and added when the predicted scores were below the mean.

---

*See Appendix B for these equations.

Another adjustment was necessary due to the finding of a consistent difference between the two scorers on the Circles Test of creativity, although the scorer agreement on the sub-scores of this test is quite high. This adjustment was made as follows: Each mean was adjusted proportionally for the overall difference in means between the two scorers. For example, the fluency scores assigned by one scorer were multiplied by .88 in order to proportionally reduce them.

In addition to the correlation matrix for all teachers, the same analysis was conducted within each of the five grades. The limitation of this approach is the small number of cases (teachers) per grade. However, it permits study of relationships which may differ from grade to grade.

In order to examine interactions between teacher characteristics and student characteristics when other student behavior variables constitutes the dependent variable, factorial analyses of variance were carried out as follows.

Within each of the five grades two teachers were selected at each of the extremes of the Factor I (warm, permissive vs. cold, controlling) variable. Thus, a total of ten at each extreme were obtained. Within each of these classes five students were selected to fit each of two groups. Group I was defined as high intelligence, low dependency and Group II was defined as low intelligence, high dependency. Bivariate scatter plots were made for each class using total raw scores on the CMM as one variate and scores on the questionnaire measure of dependency as the other variate. Selection of cases was then made. In 14 of the 20 classes no overlap on either variable was necessitated. For the other six classes some overlap was necessary though in all cases on one of the two variates only a decision had to be made whether to utilize the same cutting scores for all classes within a given grade or to select students on the basis of their positions within their class. The two procedures would not give identical results. The latter procedure was followed. There is no question, however, but what different groups were obtained within each grade as shown in Table 3.01.

This design permits the total variance for each dependent measure to be partitioned into the following components:

	d.f.	
Between Student Characteristics (Group I vs. Group II)	1	
Between Teachers	19	
Between Factor I groups (high vs. low)		1
Between grades		4
Interaction Factor I by grade		4
Residual between teachers		10
Remaining Interactions		
Student Characteristic by Teacher	19	
Student Characteristic by Factor I		1
Student Characteristic by Grade		4
Student Characteristic by Factor I by Grade		4
Residual		10
Within	160	
Total	189	

Where an interaction with grade reached or approached significance, separate analyses for each grade were made.

This analysis was repeated with a change in one major dimension. Instead of selecting high and low teachers on the basis of our Factor I behavioral dimension, they were selected on the basis of high or low scores on the questionnaire measure of control need (TPS) which had been found to predict Factor I.

TABLE 3.01

COMPARISON OF 'PUPIL CHARACTERISTIC' GROUPS WITHIN EACH GRADE-- ANALYSIS OF VARIANCE I

	HIGH INTELLIGENCE-- LOW DEPENDENCY				LOW INTELLIGENCE-- HIGH DEPENDENCY			
	INTELLIGENCE		DEPENDENCY		INTELLIGENCE		DEPENDENCY	
	$\overline{X}$	S	$\overline{X}$	S	$\overline{X}$	S	$\overline{X}$	S
Grade 1								
High Control Teachers	61.6[1]	4.0	5.4	1.0	47.2	6.5	7.8	0.4
Low Control Teachers	58.9	5.2	4.6	1.5	48.5	6.2	7.5	0.9
Grade 2								
High Control Teachers	75.0	5.2	5.9	0.8	64.0	2.3	8.0	0.6
Low Control Teachers	69.3	7.8	6.5	0.8	63.0	3.9	8.7	0.5
Grade 3								
High Control Teachers	79.9	4.7	4.5	1.9	68.1	4.0	8.4	1.0
Low Control Teachers	77.3	2.1	5.8	1.1	66.2	7.8	8.2	0.4
Grade 4								
High Control Teachers	81.6	8.3	4.5	1.2	59.1	5.0	8.4	0.9
Low Control Teachers	74.4	8.4	5.6	1.1	47.9	5.8	7.9	0.8
Grade 5								
High Control Teachers	99.4	5.3	3.4	0.9	67.6	10.8	7.1	0.7
Low Control Teachers	93.9	8.2	3.4	1.4	68.1	8.7	7.0	0.9

[1] Note.--These are raw scores. Further the CMM test level changes between grades three and four rendering comparisons of grade 3 means with grade 4 means (for example) meaningless.

CHAPTER IV

PHASE I: RESULTS

Before examining the intercorrelations among the various measures, it was necessary to determine that significant differences existed among the teachers (i.e., classes) with respect to the variables of interest. Accordingly simple analysis of variance was carried out within each grade on several of the measures of primary interest. The measures so analyzed were as follows: Gain scores in the three achievement areas, reading vocabulary, reading comprehension, and arithmetic; CMM total score, Incomplete Figures Originality, Circles Originality, Test Anxiety, Liking for School, Perceived Affiliation. On these measures significant differences at the .05 level and in many instances at the .01 or .001 level were found with the following exceptions: Reading Comprehension Gain, grades 2 and 5 (in grade 2 this reflects primarily the extremely small variance in gain across individuals which, in turn, appears to result from lack of ceiling on the post-test--the distribution of post-test scores is highly skewed); Incomplete Figures Originality, grades 1, 3, and 5; Circles Originality, grades 1, 4, and 5; Liking for School, grade 3. Thus, it appears that with respect to most of the variables of interest, significant differences among teachers within each grade do exist. The finding of significant class differences in intelligence means that this variable must be taken into account when examining other relationships. To presage the discussion to follow it may be well to indicate here that differences in class intelligence do notaccount for the important results. This is to be expected since one would not expect the teacher measures used in this study to relate to variations in class ability of the magnitude found here. Further, measures of pupil behavior based on class averages do not show consistent correlations with mean CMM.

Before proceeding further it would be well to point out the assumptions involved in the preceding statementssince in future discussion the phrase 'differences among teachers' will be used a great deal. The phrase 'among teachers' or ' between teachers' in this study refers in some cases to differences which are demonstrably differences among teachers as in the case of the teacher measures as such, i.e., the classroom observation and test data on the teachers. In other instances, however, such as those just mentioned, the phrase 'among teachers' actually refers to differences among the classes of such teachers. In the case of all but the achievement gain measures the devices administered were given once to the classes of the teachers. Since in all cases these devices were administered well into the school year, it seems legitimate to attribute at least a part of observed class differences to the impact of the teacher. It must be kept in mind, however, that any such observed difference may be attributable to other factors, such as selective placement of the students within the classes of the teachers; thus, the fact that teacher 'A' receives a test anxiety score for her pupils considerably higher than does teacher 'B' may reflect a selective placement within the classes in addition to or rather than the impact of the teacher *per se*. This problem is obviated to a large extent in the case of the achievement measures since a gain score is used which, in effect, balances out prior impact on the students within a particular class. Further, the attributing of differences in scores of the students to the impact of a particular teacher would appear to be more defensible in grade 1 since at this point the students will not have been exposed to other teachers or to the school system as such prior to their experience in this class. In addition, it is well to keep in mind the caution that even in the case of the achievement scores, since all variables which might conceivably be related to gain in achievement areas, as well as influences upon the class aside from the teacher have not been controlled, one must exercise caution in attributing observed differences to the impact of the teacher as such.

Analysis of Combined Grades Matrix

When one is contronted with approximately ten thousand correlations, some basis of systematic analysis must be established. The following procedure was followed. First, certain areas of particular interest were demarcated for study. Although the various intercorrelations among the pupil measures (class averages) are of interest, they were excluded from consideration in this part of the analysis. Thus, we are left with the relationships between the teacher measures and the pupil measures and among the teacher measures themselves. With respect to the latter, the intercorrelations among the teacher measures which have been studied previously, were only checked to ascertain (as is the case) that they follow much the same pattern as was the case in our previous study. Relationships between these measures and the OSCAR measures obtained during the most recent data collection were, however, examined. After restricting our attention to these parts of the entire matrix, the following decisions were made. First, those correlations where expectations existed were examined. Thus, in the light of our previous research, we had expected the teacher behavior measures; Factor 1, Control and Affiliation Ratings, and the TPS Control need score to be the most powerful predictors of student behavior; hence, we looked first at relationships with these variables. This phase of the procedure conforms very closely to hypothesis testing procedures and is the preferable approach. However, the magnitude of the matrix precluded the setting up of a great many specific hypotheses; hence, we were forced to a more empirical approach. First, we excluded from consideration any correlations which did not reach statistical significance at the .05 level. Next we examined the correlations from the following point of view. Each measure of teacher characteristics was examined across the various indices of pupil behavior and each pupil measure was similarly examined across the range of teacher measures. Only those variables which showed more than a chance number of significant correlations were studied further.

Another way of dealing with a matrix of this type which immediately suggests itself is through factor analysis. Before proceeding with the factor analysis, however, it seemed necessary to satisfy the condition that any relationships existing in the teacher matrix as a whole maintain themselves within at least chance departure within each of the five grades since many of the measures here employed do relate significantly and in some cases quite highly to grade. This is not true of our original teacher measures nor is it true of the achievement gain scores, but it is true of the OSCAR measures and of the other measures of pupil behavior. Even if this were not the case, it seems necessary when studying a group comprised of students selected from five grades to have some confidence that the relationships exist within each of these subsamples before submitting the overall matrix to factor analysis. This point will be discussed further later on.

Previously Obtained Teacher Measures As Related To OSCAR

In general the relationships between our previous teacher measures and the OSCAR were disappointingly low. This may in part be attributable to differences in some of the variables assessed by the two devices; that is, the three summary variables resulting from previous factor analysis of the OSCAR are not directly comparable to the measures which we had employed. However, one of these measures designated 'emotional climate' in the classroom would certainly have been expected to relate to our measures of control and affiliation but this is not the case. Some of the sub-scores on the OSCAR do relate as expected with our measures but to a lesser extent than we would have hoped; that is, the category labeled Supportive Teacher Behavior does correlate -.30 with TPS Control and .30 with the Rating Affiliation. Also, the category labeled Teacher Hostility correlates -.25 with Factor 1, which is low but in the expected direction. These correlations are contaminated by the rather high correlations between the OSCAR measures and grade level. Consequently, it is necessary to examine these relationships in more detail within each of the five grades.

Teacher Measures and Pupil Measures

Looking next at the relationships between our original measures of teacher behavior and the indices of pupil behavior, we find that our original hypotheses as to the teacher variables which would be the strong predictors are only partly correct. The ratings on Control and Affiliation, and the TPS Control score do not turn out to be very powerful predictors of measures of pupil behavior. This being the case, we are forced to give more attention to an empirical analysis of the data. Looking first at the gain scores in the achievement areas, a dimension of major interest, we find that the most effective predictor is the USB Affiliation score which correlates negatively with gain in the areas of reading vocabulary and reading comprehension, the former correlation being -.25,[1] the latter -.33. In addition, the Rating Recognition shows a positive correlation of .35 with gain in reading vocabulary. The OSCAR category 'shows affection for pupil' correlates -.36 with arithmetic gain.

---

[1] A correlation of .25 is required for significance at the .05 level. The only exceptions are the 'Consequences' measures which require .32 and the Russell Sage Test measures which require .27.

For the creativity tests the rating recognition measure shows correlations running between .32 and .44 on the various sub-tests of both the Circles and Consequences Tests. Also, the control score on the objectives of education test gives negative correlations of slightly lower magnitude, -.22 to -.32 on the same measures. The Situations Affiliation score correlates negatively with all three sub-tests of the Consequences test, correlations being .36 to .49. Finally, the Verbal Statements Achievement measure correlates negatively with both the Circles and Consequences sub-tests, the correlations ranging from .24 to .30. Once again it is necessary to consider these relationships within grade due to the rather high correlations of the creativity measures with grade. At this point, however, these data suggest that the 'stimulating' characteristic tends to foster more creative behavior on the part of the student, whereas there is some suggestion that the more affiliative teacher who selects solutions to classroom problems which would seem to be in accordance with a more affiliative approach and who is at the same time supportive of more control in education, tends to work against a more creative performance. This interpretation is extremely tentative since, as will be discussed later, only the 'Situations-Affiliation' correlation receives much support from the 'within grade' analysis.

In regard to the pupil questionnaire measures the only teacher measure to show a significant correlation with the Test Anxiety scale is the OSCAR 'Supportive Behavior' score which correlates -.44. Factor 5 correlates significantly with liking for school (.32) suggesting that the less academically oriented teacher is better liked. The measure of pupil affiliation need correlates .31 with Situations Affiliation and .36 with Situations Recognition and -.34 with Verbal Statements Affiliation. The score intended to reflect perception of the teacher as controlling correlates .32 with Objectives Recognition, .29 with Objectives Control, and -.34 with the stimulating dimension, Factor 2. These latter results do not provide a coherent picture.

On the Russell Sage Test there are some rather interesting relationships with teacher behavior. First, the rating on Achievement orientation of the teacher correlates positively with the adequacy of the planning measure (.32) and with the measure of success on one of the two problems (.36) suggesting that the teacher who emphasizes achievement and doing a good job tends to get a higher level of planning from the children and possibly a higher level of performance on the solution of the task although this obtains for only the first of the two problems. This same measure (achievement orientation of the teacher) correlates negatively (-.31) with social behavior during the operations phase of the problem for the second problem only. These results taken together suggest that the teacher who emphasizes achievement orientation or doing a good job fosters adequacy of planning which in itself may be taken as an indication of achievement and also instills the desire to do a good job on the task, but that when the task itself requires social interaction and cooperation that this behavior on the part of the teacher tends to foster an attitude in which social cooperation breaks down, resulting in greater tendency toward social disruption. Some additional support for this point of view is provided by the TPS achievement score which correlates negatively (-.38) with the measure of overall atmosphere during the operations phase. This correlation tends to be somewhat higher on the first of the two problems but is maintained across both problems. The correlation between 'atmosphere' and Factor 3 (-.35) suggests that teachers low in ego strength, as rated, foster a less desirable atmosphere during the operations phase.

The 'supportive behavior' score on the OSCAR reveals suggestive positive correlations with operations atmosphere and involvement (.31 and .41) and negative correlations with adequacy of planning, ideas, etc. (-.24 to -.40). Since both the Russell Sage planning measures and this OSCAR score are significantly related to grade, partial correlations were computed which reduce the magnitude of each negative correlation by approximately .10.

Analysis of Teacher Matrices by Grade

In this analysis intercorrelations among the measures were analyzed within each of the five grades. The procedure followed was similar to that followed with the combined teacher matrix differing only in the level of significance which was required. It was decided that correlations significant at the .10 level within each grade would be retained for further study. The number of teachers per grade varies from ten to fifteen.

Previously Obtained Teacher Measures As Related To OSCAR

The within grades analysis is consistent with the combined grades analysis in showing a pattern of general agreement among logically similar measures but at a much lower level of correspondence than was hoped. The limited number of OSCAR observations may account in part for this finding. Table 4.01 shows the correlations between Factor 1 and the OSCAR scores logically expected to be related most highly. In grades 2 through 5 there appears to be a correspondence in that high scores on Factor 1 are negatively associated with teacher hostility (verbal and non-verbal) and positively associated with supportive teacher behavior (primarily non-verbal). In grade 1, however, the only sizable relationship is a negative correlation with 'supportive verbal behavior,' an unexpected result which indicates that the most warm, permissive teachers as viewed by the observers actually bestow less praise and encouragement than those viewed as more aloof and controlling.

Teacher Measures and Pupil Measures

The first domain of pupil behavior to be explored is that of achievement gain. The first question to be answered is whether the number of significant correlations exceeds that to be expected by chance; and

TABLE 4.01

CORRELATIONS BETWEEN FACTOR 1 AND SELECTED OSCAR MEASURES ACROSS GRADES

OSCAR MEASURE	GRADE 1	2	3	4	5
Teacher Hostility	-.08	-.26	-.46	-.34	-.12
Hostile Teacher Statements	-.03	-.45	-.80	-.31	.04
Supportive Teacher Behavior	-.29	-.11	.30	.59	.63
Supportive Teacher Statements	-.68	-.16	-.05	.38	-.11
Teacher Shows Affection	.19	-.26	.24	-.15	.66
Supportive Non-Verbal Behavior	.06	.32	-.05	.74	.65
Problem Structuring Behavior	.20	.33	.03	-.15	.06

the answer to this question is clearly that it does. Out of the total of 435 correlations one would expect 43 or 44 to be significant at the .10 level by chance, whereas the number actually significant at this level is 103. Furthermore, in certain grade-area combinations, notably in grade 2, reading vocabulary, and in grade 3, arithmetic, the number of significant correlations is very nearly half the correlations. We are, therefore, encouraged to believe that our measures of teacher behavior and teacher characteristics are related to measures of achievement gain at more than a chance level of significance. From this point our analysis took two directions. The first was to seek relationships which might be consistent across all five grades or which seemed to show a pattern of shift across the five grades. Thus, any teacher measure which provided significant correlations in two grades within an achievement area was further pursued in that all five correlations across the five grades were studied. The result of this investigation suggests that there is only one relationship which maintains itself across all five grades and that is the relationship between gain in reading vocabulary and the observers' rating of 'Recognition' on the part of the teacher. The nature of this correlation suggests that the more stimulating teacher who is rated as higher in recognition-type behavior obtains a greater growth in reading vocabulary regardless of grade. The lowest correlation obtained is .29 in grade 4 ranging up to .65 in grade 1. The correlations observed on the combined teacher matrix between USB affiliation and the two reading gain measures are attributable almost entirely to grade 1 wherein there is a high negative correlation between affiliation need on the USB and reading vocabulary gain and comprehension gain, the correlations being -.74 and -.55 respectively. However, in the other four grades on both measures the correlations drop very close to zero and exhibit no trend.

Factor 1, which we had expected to be one of our better predictors of student change, shows a very interesting pattern with respect to gains in reading comprehension but one which is not consistent. Four of the five correlations are of substantial magnitude, but there is no consistent pattern across grades. The correlations in order from grade 1 through 5 are: -.66, .59, -.53, -.12, and .55. It appears somewhat strained to attribute this series of correlations all but one of which are significant to simply chance fluctuation and yet it makes little psychological sense. There appear to be no other important trends or consistencies in individual predictors of achievement gain. It may, however, be worth noting that the four need scales on the USB do appear to show a patterning which carries across all four needs, to the effect that in the first grade high needs in all four areas are significantly negatively correlated with achievement gain. The correlations in grade 2 are once again negative but lower, whereas the correlations begin to shift from negative to positive in grade 3, and in grades 4 and 5 the correlations are all positive suggesting that a high level of need in general is deleterious to achievement gains in the earliest grades but fosters such gain in the upper grades.

Looking at the relationships between the OSCAR measures and achievement gain there is very little which emerges. For the arithmetic gain there is a suggestive negative relationship with 'affectionate behavior' on the part of the teacher in the early grades. The correlations by grade are -.64, -.61, -.22, -.12 (the grade 5 correlation is meaningless since it is based on only 5 teachers).

Within grade correlations between the creativity scores and measures of teacher behavior indicate only one promising predictor and that is the situations affiliation score which tends to correlate negatively with scores on the consequences test, which was employed in grades 3, 4, and 5. The magnitude of the negative correlations is quite substantial in grade 3, somewhat lower in grade 5 and non-significant in grade 4. Nevertheless, there does appear to be some consistency in this relationship. This is not true of any of the other measures of teacher behavior although once again there are many more than the chance number of significant correlations at the 10 percent level. They do not, however, reflect any kind of meaningful pattern. On this point it should be noted that one could not have expected very promising results from this particular analysis since presumably the child's score on the creativity test will be influenced by many other variables which were not controlled since the score used was not a gain score across the period of time the child spent with this particular teacher but simply a status score obtained toward the middle of the school year. The major purpose of collecting these scores was to investigate their relationship with other pupil measures so as to provide a more knowledgeable approach to changes in creativity score which was conducted in Phase II of this study.

Of the pupil questionnaire measures two were considered to be particularly appropriate as reflections of teacher behavior, namely the 'test anxiety' scale and the 'liking for school' scale which in terms of the items contained therein might more appropriately be called a 'liking for teacher' scale. The other questionnaire measures are either intended to reflect more basic personality attributes of the children, such as dependency need or reflect characteristics of perception of school which might be considered to be an elaboration of the liking for school score.

With regard to liking for school one relationship seems to maintain itself across all five grades and that is a positive correlation between liking for school and little concern with academic matters on the part of the teacher (Factor 5). Thus, it appears that the better liked teachers are those who are less oriented toward strictly academic achievement. The correlation is .47 and .53 in grades 2 and 3 respectively and somewhat lower in the other three grades (.21, .06, and .15) but positive in all cases. The next set of relationships which is of particular interest pertains to Factor 1 which in terms of prior expectation was hoped to be one of the stronger predictors. There is a correlation of -.51 in grade 1, -.19 in grade 2, then shifting to the positive .52 in grade 3, .24 in grade 4, and .58 in grade 5. The fact that three out of the five correlations are significant and large combined with the apparent trend strongly suggests that in the first grade and to a lesser extent in the second, it is the more controlling, somewhat aloof teacher who is better liked, but that this pattern shifts around second to third grade, and that thereafter it is the warm, more permissive teacher who is better liked. There appear to be no important relationships with the OSCAR measures.

Looking at relationships with test anxiety there is one very clear relationship which seems to emerge and that is a negative correlation with the OSCAR measure of "supportive teacher behavior" indicating that the more supportive teacher does not foster as high a degree of test anxiety. The correlations range in magnitude from -.10 in grade 2 to -.67 in grade 3. Hence, it appears that one can have some confidence in this generalization across all five grades. A very similar relationship is found with one of the sub-scores making up the overall score on supportive behavior, this being the non-verbal supportive behavior on the part of the teacher. There is also the suggestion that in the higher four grades problem structuring behavior on the part of the teacher fosters test anxiety as might be expected, positive correlations existing in grade 2 through 5 (.35, .27, .38, .56) but not in grade 1 where the correlation is negative (-.25) and again this tends to support the previous argument that control on the part of the teacher tends to have a negative effect on the upper grades, but that this is not the case in the first grade where emphasis on control and problem structuring on the part of the teacher seems to have a positive effect.

There appear to be no appreciable correlations with teacher measures other than OSCAR which maintain themselves across all five grades. There is, however, an interesting trend which maintains itself across two indices of teacher behavior, namely the ratings on recognition and the Q-sort Factor 2 which is a recognition-type factor, being defined as quiet, dull vs. stimulating. In the first grade the relationship between test anxiety and recognition or stimulating behavior appears to be negative suggesting that the more active, more stimulating teacher generates less anxiety in the first grade. This pattern tends to hold although the correlations are of lesser magnitude in grades 2 and 3 but by grade 4 the pattern has shifted such that the reverse relationship is the case. The Factor 2 correlations are .44, .09, .29, -.39, -.39. The teacher who is stimulating seems to generate more anxiety on the part of the student at the higher grade level.

On the Russell Sage Test there are three measures of interest. The first relates to success on the problems. On this measure there seems to be no systematic relationship across grades. The positive correlation between success and the rating of the teacher as achievement oreinted which was observed in the combined-grades matrix holds up for the first three of the grades; that is, grades 2, 3, and 4 but not in grade 5 where it is highly negative (.54, .39, .32, -.70). This pertains to the first problem only. The measure of overall success on the problem, however, does seem related to one of our major dimensions, in that success on the problems combined is positively correlated with Factor 1 across all four grades (.21, .40, .12, .33), and supporting this are negative correlations with the ratings on control in all four grades (-.46, -.22, -.30, -.48) and additional partial support on the Ratings Affiliation where positive correlations obtain for three of the four grades (.36, .39, -.03, .39). This issue is somewhat complicated by the fact that in only one of the grades is the correlation between success on problems one and two significant (-.07, .31, .05, .62). The pattern emerging here then is that the teacher who is more permissive and less controlling would appear to foster better problem solution on this particular test.

The second set of scores of interest pertains to the adequacy of the planning stage on this test and here we find that there is a consistent pattern with respect to the ratings on achievement score. Thus, on three of the scores (adequacy of communication, adequacy of ideas proposed, and adequacy of the final plan which eventuates) there are significant positive correlations in several of the grades and a pattern of positive correlations across grades indicating that teachers who are viewed as more achievement oriented tend to have classes who plan more adequately on this task (see Table 4.02). The correlations are about the same whether one looks at problem one alone, problem two alone, or the combined planning score.

The third area of interest pertains to the behavior of the group during the operations phase or when the problem is actually being worked on. If one looks at the combined scores based on both problems, there is little by way of relationship with the teacher variables. Looking just at the second problem, however, a pattern of relationships with the Achievement Rating again emerges. Thus, correlations across grades are, with one exception, all negative and in several instances significant for the scores indicating congeniality of atmosphere (see Table 4.02) and for scores indicating lack of noise and absence of hostility. The nature of these correlations would indicate that the classes with the more highly achievement oreinted teachers tend to have a somewhat noisier, more hostile atmosphere. In summary these findings suggest that the teacher who is more achievement oriented does communicate this to her students and that this affects the adequacy of the planning of a task such as this, resulting in higher level of attack on the problem and problem solving skills. However, these same students, when involved in the problem, tend to break down in terms of social interaction becoming more agitated and hostile, although this does not seem to emerge particularly until the second problem is attempted. It also appears that teachers viewed as higher in ego strength foster a more congenial atmosphere.

Looking next at the OSCAR measures, on the planning phase the only variable which appears at all promising pertains to the communications phase of planning where negative relationships between adequacy and 'overt hostility' on the part of the teacher appear to exist. On the operations phase (problem two only) where there are some promising relationships with the other teacher measures, one dimension shows up fairly clearly across all three indexes of pupil behavior: 'general atmosphere,' 'lack of noise,' and 'absence of hostility.' In grades 2 through 4 there are rather high positive correlations between the general desirability of the group atmosphere; i.e., lack of hostility, lack of noise and generally friendly atmosphere and the extent of supportive teacher behavior, suggesting that the teacher who is more supportive has an impact on group structure such that it is less likely to break down under this kind of task.

Summary Table. Those relationships which have emerged as most promising are summarized in Table 4.02.

TABLE 4.02

SUMMARY OF THE MOST PROMISING RELATIONSHIPS BETWEEN TEACHER MEASURES AND PUPIL MEASURES

PUPIL MEASURE	TEACHER MEASURE	CORRELATIONS BY GRADE 1	2	3	4	5	1-5
Reading Vocabulary Gain	Rating Recognition	.65**	.35**	.44*	.29	.56*	.35***
Reading Comprehension Gain	Factor 1 (High score = Warm, Low Control)	-.66**	.59**	-.53*	-.12	.55*	.04
Arithmetic Gain	OSCAR - Demonstrates Affection	-.64**	-.61**	-.22	-.12	-.95**	-.36***
Test Anxiety	Rating Recognition	-.47*	-.37*	.29	.19	.05	.11
Test Anxiety	Factor 2 (High score = Quiet, Dull)	.44*	.09	.29	-.39*	-.39	-.15
Test Anxiety	OSCAR - Supportive Behavior	-.43*	-.10	-.67**	-.31	-.61*	-.44***
Test Anxiety	OSCAR - Problem Structuring	-.25	.35*	.27	.38*	.56*	.06
Liking for School	Factor 5 (High score = Little Concern with Achievement	.21	.47*	.53*	.06	.15	.32**
Liking for School	Factor 1 (High score = Warm, Low Control)	-.51*	-.19	.52*	.24	.58*	.14
Russell Sage Planning--Communication	Rating Achievement	----	.41*	.12	.44*	-.18	.21**
Russell Sage Planning--Communication	OSCAR Hostility	----	-.20	-.46*	-.09	-.88**	-.25**
Russell Sage Planning--Ideas	Rating Achievement	----	.19	.36*	.38*	.33	.29**
Russell Sage Planning--Plan	Rating Achievement	----	.37*	.34	.52*	.07	.32**
Russell Sage Operations--Atmosphere Problem 2	Rating Achievement	----	-.25	.10	-.35	-.91**	-.31**
Russell Sage Operations--Atmosphere Problem 2	OSCAR Supportive Behavior	----	.45*	.66**	.52*	.04	.24**
Russell Sage Operations--Atmosphere Problem 2	TPS Achievement	----	-.47*	-.29	-.31	-.56*	-.38***
Russell Sage Operations--Atmosphere Problem 2	Factor 3 (High score = High Ego Strength	----	.43*	.13	.49*	-.08	.27**
Russell Sage Success	Factor 1 (High score = Warm, Low Control)	----	.21	.40*	.12	.33	.24**
Russell Sage Success	Rating Control	----	-.46*	-.22	-.30	-.48*	-.28**

* Sig. .10 level　　** Sig. .05 level　　*** Sig. .01 level

Note.--The N per grade is approximately 12.

Analysis of Correlations Within Grade 1

In addition to exploring consistencies or patterns of correlations across the five grades, it is possible to examine the set of correlations within each grade for meaningful relationships. With N's as small as ten to fifteen within each grade such a pursuit is certainly tenuous. Nevertheless, certain patterns seem to emerge within the first grade which warrant discussion. Three of the correlations within grade 1 previously mentioned are of particular interest. First, the negative correlation of -.66 between Factor 1 and Reading Comprehension gain suggesting that the more controlling, less permissive teacher achieves a greater gain. The correlations with Reading Vocabulary and Arithmetic gains are in the same direction though not significant (-.36 and -.30). Second, the correlation, again negative between the USB Affiliation score and Reading Comprehension gain (-.55) supporting the same interpretation, and finally the negative correlation between Liking for School and Factor 1 (-.51) suggesting that the better liked teacher within this grade was the more controlling, less permissive teacher. The lowest of these correlations is significant at the .10 level and the highest at the .02 level. Pursuing this further to determine whether the remaining intercorrelations among these four measures would add further consistency to this interpretation indicates that two of the three provide support in that USB Affiliation correlates positively with Factor 1, as would be expected (.39), and Liking for School correlates as would be expected with gain in Reading Comprehension (.38). Although these values are not high, they are in the predicted direction. The remaining correlation (USB Affiliation and Liking for School) however, does not support this interpretation being .00. These correlations are summarized in Table 4.03.

There is a complication in interperting these data in that the first grade is the one grade of the five wherein the rating measures of teacher behavior and the Q-sorts which are both based on observer impressions are not consistent within themselves nor do they receive support from the more objective measure of teacher statements. We had found in our earlier study that there was good consistency between the Rating Affiliation and Rating Control scores and the Factor 1 score and also considerable support from the Verbal Statements-Control measure. This holds up in all of the four higher grades when analyzed individually but does not hold up very well in the first grade (see Table 4.04). The correlation between Verbal Statements-Control and Factor 1 is low though in the expected direction. The correlation between Verbal Statements-Control and the Rating on Control is in fact negative. The ratings, however, are not related to measures of pupil behavior; nevertheless this finding raises a question as to the meaning of the observational measures in this grade. Our tentative interpretation is that the inclusion of reprimands of our control classification of teacher statements may be responsible. It appears that in the upper grades those teachers who reprimand most were also judged as most controlling but that in the first grade this relationship is reversed. We intend to clarify this point by re-classifying the teacher statements to exclude reprimands from the control category.

Additional evidence is provided by the results for grade 1 on the OSCAR measures of teacher behavior which were obtained in a different year and completely independent from the previous teacher measures. Here we find that the Liking for School measure is positively correlated with the category of 'supportive verbal behavior' on the part of the teacher (.58) but negatively correlated with the score reflecting overt teacher affection for the pupil (-.51). Further, liking for school is negatively correlated both with hostile verbal behavior (-.63) on the part of the teacher and the category reflecting hostile teacher behavior toward the pupil (-.69). Further, the gain scores in Reading Vocabulary and Comprehension are negatively correlated with 'hostile statements' on the part of the teacher (-.36 and -.53) and positively correlated with 'supportive verbal behavior' (.46 and .41). Thus, it would appear that although hostile behavior on the part of the teacher is certainly not to be desired, that overt affection on the teacher's part is not liked by the pupils and that supportive verbal behavior (which in large measure reflects praise and encouragement) is effective in promoting both liking for the teacher and gain in achievement.

One of the most striking aspects of these results is the seemingly inconsistent relationship between observer impressions (Factor 1) and the OSCAR 'supportive statements' score. Each of these is highly related both to pupil liking for school and achievement gain in reading. Yet the correlation between them is the reverse of what was expected since the teachers scoring high on 'supportive statements' are rated as less warm and permissive (-.68).

Our interpretation is as follows: Overt hostility on the part of the teacher appears undesirable as would be expected. Further, verbal support from the teacher appears to be quite desirable in this grade though overt affection appears not to be conducive to liking for the teacher. A classroom which is less permissive, i.e., more highly controlled by the teacher, appears desirable. Thus, our tentative picture of the ideal first grade teacher is one whose affiliation needs are not excessive and who engages in little overt affectionate behavior. She would provide a good deal of structure, i.e., be quite controlling but at the same time very free with praise and encouragement. Clearly this hypothesis needs further testing.

Relationships Between Observer Perceptions and Pupil Perceptions

One of the hypotheses of the study was that some agreement would be found between the observations on the part of the pupils and the observers under the conditions of the present study. There is some support for this hypothesis as one progresses to the higher grades, but not in the earliest grades. Thus, in grade 1, as shown in Table 4.05, the pupil's perception of the teacher as affiliative is only slightly related to the observer rating of affiliation and is related in a negative direction to the rating on control; that is, the teachers viewed by the observers as the most controlling are perceived by the pupils as warmer which is again in accord with our previous discussion; and Factor 1 which

TABLE 4.03

INTERCORRELATIONS AMONG MEASURES FOR GRADE 1

	FACTOR 1	USB AFFILIATION	LIKING FOR SCHOOL
Gain in Reading Comprehension	-.66	-.55	.38
Factor 1		.39	-.51
USB Affiliation			.00

TABLE 4.04

INTERCORRELATIONS AMONG MEASURES OF TEACHER BEHAVIOR ACROSS THE FIVE GRADES

	GRADE 1				GRADE 2				GRADE 3		
	T.S. Con.	Rat. Con.	Rat. Aff.		T.S. Con.	Rat. Con.	Rat. Aff.		T.S. Con.	Rat. Con.	Rat. Aff.
Factor 1	-.30	.00	.15	Factor 1	-.39	-.47	.56	Factor 1	-.73	-.74	.80
T.S. Con.		-.46	-.31	T.S. Con.		.62	-.54	T.S. Con.		.38	-.54
Rat. Con.			.18	Rat. Con.			-.63	Rat. Con.			-.85

	GRADE 4				GRADE 5		
	T.S. Con.	Rat. Con.	Rat. Aff.		T.S. Con.	Rat. Con.	Rat. Aff.
Factor 1	-.59	-.47	.46	Factor 1	.11	-.51	.60
T.S. Con.		.31	-.51	T.S. Con.		.00	.04
Rat. Con.			-.42	Rat. Con.			-.86

is a combination of the observer perceptions on affiliation and control is negatively correlated though low. In grade 2 the relationships are all extremely low indicating very little correspondence between the pupil perception and the observer perceptions. Beginning in grade 3, however, the correspondence begins to increase and from there on most of the correlations are in the expected direction and do tend to form a pattern. If one looks at perceived affiliation in relationship to Factor 1, there is a rather strong agreement with observer ratings in grades 3, 4, and 5, though it should be noted that Factor 1, which at the high end reflects both low control and high affiliation as judged by observers is related primarily to the student's perception of affiliation; that is, these teachers are seen as warmer in the upper three grades but not necessarily as less controlling. The preceding discussion rests, of course, on the premise that our pupil questionnaire measures of perceived control and affiliation adequately measure these variables.

Results of Factorial Analyses

The first series of factorial analyses used the Factor 1 score as the basis of separation of teachers into two groups for one main affect and used as a second main affect a combination of intelligence scores and dependency scores on the part of the students. Four teachers, two each at the extremes, were utilized in each grade making a total of 20 teachers in this analysis. The first analysis to be performed included all 20 teachers and hence included a between grades effect as one main affect. In the event that this analysis seemed promising and particularly if there were an interaction of some other variables with grade, then separate analyses were performed for each of the five grades on the particular dependent variable in question. The principle hypothesis of interest in these analyses pertains to the expected interaction between the Factor 1 measure on the part of the teacher and the student characteristics, it being hypothesized initially that students who were high in intelligence and low in dependency would "do well," (whatever that may mean, on each variable) with a teacher who was low on Factor 1; that is, the more controlling, less affiliative teacher whereas the student who was low on intelligence, high in dependency need would show the reverse effect and would "do better" with the warmer, less controlling teacher.

These analyses utilizing the creativity scores as the dependent variable provide very little aside from what appear to be random significance of effects and hence are discussed no further. With respect to the Reading Vocabulary score a significant interaction is suggested. Although the F value is not significant at the .05 level, it does approach the necessary value. In addition, there is a significant between grades effect which indicates only that the four teachers selected in a particular grade achieved somewhat better results than the four teachers so selected in another grade. This does not reflect a difference in achievement between grades overall since this effect was ruled out by the nature of obtaining gain scores. When this analysis was repeated within each of the five grades, it became evident that most of the effect observed was attributable to the first grade wherein the interaction value is significant at about the .12 level but in the opposite direction from the hypothesis. Thus, the more dependent students showed greater gain with the less permissive teachers and the less dependent students gained more with the more permissive teachers. This finding is, however, consistent with our present interpretation that the first grader has a great need for structure in the classroom. Thus, we find more gain under the more controlling teacher (correlation analysis) and especially for the more dependent of the students (factorial analysis). None of the main affects or interactions are significant in any of the other four grades.

On the Reading Comprehension score there are no significant effects other than a significant main effect for between students characteristic. Inasmuch the intelligence dimension for selecting students should not have a bearing when the gain scores are used since as the gain score itself is controlled for intelligence, the gain score should in this analysis reflect differences between students with respect to the dependency dimension. However, since for the entire group in grade 1 there is no significant correlation between gain in reading comprehension and dependency need, this particular significant F would appear to be an artifact of the particular sample employed.

In light of the correlation previously discussed between Factor 1 and gain in Reading Comprehension for grade 1, one would have expected this main affect to be significant. Upon inspection of the data it is apparent that although the two teachers high in Factor 1 do achieve less gain than the two teachers low in Factor 1, two of the teachers, one from each group, are sufficiently similar to account for the fact that the differences in these groups are not significant inspite of the overall positive correlation between gain and Factor 1.

Looking next at the gain in arithmetic, there are two significant main affects with the control dimension in grades 1 and grades 3: The more controlling teacher in this analysis achieves greater gain than the less controlling teacher. However, once again this is probably best regarded as an artifact since the correlations between Factor 1 and Arithmetic Gain across all teachers in each of these grades are not significantly different from zero. For the fifth grade there is an interaction between student characteristic and teacher characteristic which is in accordance with the original hypothesis and results primarily from the interaction in one particular cell wherein the two least controlling teachers achieve a significantly lower gain in arithmetic from the children who are classified as lower in dependency need.

The Sarason Test Anxiety measure as a dependent variable reveals significant effects for the between student characteristics; between teachers and between grades effects which were to be expected based on the overall relationships between these variables and Test Anxiety. The Liking for School measure shows differences only for 'between teachers' and 'between grades.'

TABLE 4.05

RELATIONSHIPS BETWEEN OBSERVER PERCEPTIONS AND PUPIL PERCEPTIONS OF TEACHERS ACROSS GRADES[1]

		OBSERVER PERCEPTIONS		
		FACTOR 1	RATING CONTROL	RATING AFFILIATION
Student Perception	Perceived Affiliation	-.28	.62	.29
		-.29	.16	-.08
		.51	-.43	.27
		.31	.07	-.06
		.73	.01	.37
	Perceived Control	-.16	-.01	-.57
		-.06	-.27	.17
		.09	.02	-.13
		-.23	.20	-.20
		-.18	.35	-.59

[1] The sequence of grades in each case is Grade 1, 2, 3, 4, 5.

TABLE 4.06

ANALYSIS OF VARIANCE I--READING VOCABULARY GAIN

SOURCE OF VARIANCE	S.S.	d.f.	MEAN SQUARE	F	
Between Student Characteristic	30.42	1	30.42	<1.00	
Between Teachers	530.60	19	27.93	1.10	
Between Control	.02	1	.02	<1.00	
Between Grades	369.80	4	92.45	3.65	(p<.01)
Interaction Control x Grade	14.68	4	3.67	<1.00	
Residual Between Teachers	146.10	10	14.61	<1.00	
Interaction--Student Characteristic by Teacher	522.98	19			
Interaction St. Char. by Control	62.72	1	62.72	2.48	(p≈.12)
Interaction St. Char. by Grade	63.28	4	15.82	<1.00	
Interaction St. Char. by Grade by Control	89.50	4	22.38	<1.00	
Residual Between Cells	307.48	10	30.75	<1.00	
Within Cells	4,052.00	160	25.32		
Total	5,136.00	189			

TABLE 4.07

ANALYSIS OF VARIANCE I--READING VOCABULARY GAIN--GRADE 1

SOURCE OF VARIANCE	S.S.	d.f.		MEAN SQUARE	F	P
Between Student Characteristic	30.63	1		30.63	<1.00	
Between Teachers	91.28	3		30.43	<1.00	
Between Control	3.03		1	3.03	<1.00	
Residual Between Teachers	88.25		2	44.12	1.01	
Interaction: Student Characteristics by Teacher	246.27	3		80.09	1.84	
Interaction St. Char by Control	119.02		1	119.02	2.72	(p≤.11)
Residual Between Cells	121.25		2	60.62	1.39	
Within Cells	1,395.20	32				
Total	1,757.38	39				

TABLE 4.08

ANALYSIS OF VARIANCE II--READING VOCABULARY GAIN--GRADE 1

SOURCE OF VARIANCE	S.S.	d.f.	MEAN SQUARE	F	p
Between Student Characteristics	14.40	1	14.40	<1.00	
Between Teachers	688.50	3	229.50	3.69	
Between Control	270.40	1	270.40	4.35	<.05
Residual Between Teacher	418.10	2	209.05	3.36	≈.06
Interaction: Student Characteristics by Teacher	750.60	3	250.20	4.02	<.05
Interaction St. Char. by Control	240.10	1	240.10	3.86	≈.06
Residual Between Cells	510.50	2	255.25	4.10	<.05
Within Cells	1,990.40	32	62.20		
Total	3,443.90	39			

TABLE 4.09

ANALYSIS OF VARIANCE II--READING VOCABULARY GAIN--GRADE 2

SOURCE OF VARIANCE	S.S.	d.f.		MEAN SQUARE	F	P
Between Student Characteristics	.10	1		.10	<1.00	
Between Teachers	56.40	3		18.80	<1.00	
Between Control	10.00		1	10.00	<1.00	
Residual Between Teachers	46.40		2	23.20	1.17	
Interaction: St. Char. by Teacher	159.50	3		53.17	2.69	
Interaction: St. Char. by Control	84.10		1	84.10	4.25	<.05
Residual Between Cells	75.40		2	37.70	1.90	
Within Cells	633.00	32		19.78		
Total	849.00	39				

TABLE 4.10

ANALYSIS OF VARIANCE II--ARITHMETIC GAIN--GRADE 1

SOURCE OF VARIANCE	S.S.	d.f.	MEAN SQUARE	F	p
Between Student Characteristics	90.00	1	90.00	1.28	
Between Teachers	1,153.20	3	384.40	5.46	
Between Control	250.00	1	250.00	3.56	≈.08
Residual Between Teachers	903.20	2	451.60	6.41	<.01
Interaction: St. Char. by Teacher	296.00	3	98.66	1.40	
Interaction: St. Char. by Control	270.40	1	270.40	3.84	≈.06
Residual Between Cells	25.60	2	12.80	<1.00	
Within Cells	2,254.40	32	70.45		
Total	3,793.60	39			

The second series of factorial analyses were conducted the same as the first series except that the teachers were sub-divided into high and low control teachers on the basis of their TPS control score rather than on the basis of our observational data. Once again the analyses for the Creativity tests contributed nothing and are not discussed further. With respect to the achievement areas, the Reading Comprehension gain again shows nothing, but the Reading Vocabulary gain provides results which are consistent with those obtained in the first series of analyses in that in both the first and second grades the children classified as less dependent do better with the less controlling teacher and the children classified as more dependent do better with the more highly controlling teacher as assigned on the basis of the TPS. In the first grade the effect is almost entirely due to one cell wherein the less dependent children do much more poorly with the highly controlling teacher. In grade 2 the effect is more balanced across all four cells. With respect to gains in arithmetic the interaction in grade one is significant and in the direction consistent with prior results, in that, in particular the more dependent children did much more poorly with the low controlling teachers, this cell contributing almost all the interaction variance. There is also a significant interaction in grade 5 which is consistent with the results obtained in Factorial Analysis I and suggests that in the fifth grade the children who were low in dependence did significantly poorer with the low controlling teacher. The questionnaire measures of anxiety, liking for school, etc. produced nothing by way of interaction with this measure of teacher characteristics.

Correlations Among Gains Achieved by Teachers in the Three Achievement Areas

Although not of direct bearing on the major questions of interest in this study the correlations among achievement gains (by class) within the five grades are of interest and are shown in Table 4.11. The correlations among these measures correlated across individual pupils within each grade are shown in Table 4.12. Individual differences in pupil gains--with pre-test and CMM intelligence partialed out--are clearly demonstrated in Table 4.12. Reading comprehension gains are quite consistently correlated with gains in the other two areas whereas Reading Vocabulary and Arithmetic show less consistent but lower correlations.

With respect to teacher (class) differences, it appears that some teachers accomplish greater gains than others across all three areas. The low correlations in grade 2 are a direct result of the non-significant variability among classes on Reading Comprehension and the barely significant difference on Reading Vocabulary.

It is illuminating to compare the obtained correlations in Table 4.11 with the estimate of maximum correlation possible in light of the reliabilities of the measures themselves. Reliability estimates were obtained using the formula $r_{11} = 1-\frac{\sigma_e^2}{\sigma^2}$ where σ_e^2 is the variance of errors associated with each score. Since in this case each score is a mean gain score the appropriate error is the variance error of the mean. This value was obtained by using the 'within' mean square from the

TABLE 4.11

CORRELATIONS AMONG ACHIEVEMENT GAINS (BY CLASS) ACROSS THE FIVE GRADES

	READING VOCABULARY-- READING COMPREHENSION		READING VOCABULARY-- ARITHMETIC		READING COMPREHENSION-- ARITHMETIC		
Grade 1	.63**	(.70)	.02	(.74)	.52*	(.87)	
Grade 2	.00	(.35)	-.05	(.50)	.37*	(.40)	[.16]
Grade 3	.63**	(.73)	.19	(.77)	.26	(.84)	[.52]
Grade 4	[illegible]***	(.73)	.31	(.73)	.25	(.80)	[-.07]
Grade 5	.28	(.48)	-.14	(.77)	.16	(.52)	
All Grades	.55****		.11		.39***		

* Sig. .10
** Sig. .05
*** Sig. .01
**** Sig. .001

TABLE 4.12

CORRELATIONS AMONG ACHIEVEMENT GAINS (BY INDIVIDUAL) ACROSS THE FIVE GRADES

	READING VOCABULARY--READING COMPREHENSION	READING VOCABULARY--ARITHMETIC	READING COMPREHENSION--ARITHMETIC
Grade 1	.49	.28	.30
Grade 2	.34	.14[a]	.31
Grade 3	.34	.27	.26
Grade 4	.33	.19	.31
Grade 5	.34	.09[b]	.27

NOTE.--All r's are significant at the .001 level except 'a' which reaches the .01 level and 'b' which does not reach .05.

TABLE 4.13

ESTIMATED RELIABILITIES OF CLASS GAIN SCORES
BY GRADE AND CONTENT AREA

GRADE	READING VOCABULARY	READING COMPREHENSION	ARITHMETIC
1	.60	.82	.92
2	.44	.28	.56
3	.67	.79	.89
4	.65	.78	.83
5	.71	.32	.84

Analysis of Variance within each grade as the estimate of population variance and dividing by the average N per class within each grade (25-29) to obtain σ_e^2. The variance among scores (σ^2) is simply the variance among the means for each grade. These reliabilities are shown in Table 4.13. The estimates of maximum correlation possible are shown in parentheses in Table 4.11. These comparisons indicate that, within the ability of our gain scores to discriminate, there is a close correspondence between gains achieved in the two reading areas, as would be expected. There is further, considerable correspondence between Reading Comprehension and Arithmetic but not between Reading Vocabulary and Arithmetic.

The possibility that these correlations are an artifact of class differences in ability was checked. Although there are several correlations significant at the .10 level between class CMM and gain, the pattern is not consistent, there being both positive and negative correlations. Partial correlation, controlling CMM, has the effect of changing the correlations between Reading Vocabulary and Reading Comprehension by less than .05. The generally inconsistent relationship between Reading Vocabulary and Arithmetic is similarly, affected only slightly. For Reading Comprehension and Arithmetic, however, the pattern of low positive correlations is disrupted, in that the values for grades 2, 3, and 4 change to those shown in brackets in Table 4.11. Thus, the relationship between these two gains is unclear. The values across grade 1 remain unchanged due to extremely low correlations between any of the class gain scores and intelligence.

We conclude that some teachers do a better job of teaching reading than others and that the differences are not attributable to differences in ability level of the class. Further, within the first grade, the teacher who fosters greater comprehension in reading seems to foster greater gain in arithmetic fundamentals. There seems to be little correspondence between achieved gain (by class) in the reading skills and arithmetic skills (except for grade 1) when class ability level is controlled.

CHAPTER V

PHASE II: PROCEDURE

Phase II of the investigation may be divided into three parts as follows:

1. Evidence was collected to determine the effects of teacher classroom control on pupil gains on measures of creative thinking.

2. Classroom behavior data was obtained for groups of high and low creative pupils as measured by tests in classrooms of high and low controlling teachers. Special attention was given to possible interactions between pupil measured creativity and teacher classroom control as determiners of pupil classroom behavior.

3. Data was collected on the reliability of the Torrance creativity tests.

Selection of High and Low Controlling Teachers

Due to the practical difficulties in obtaining test data in the first grade, this phase of the study was limited to grades two through five. Within each of these grades two high control and two low control teachers were selected. The basis of selection was scores on four indexes: Factor 1, Rating Control, Verbal Statements Control, and TPS Control.

In selecting high and low controlling teachers from the original sample of 54 teachers in grades two through five, the four selector scores were converted to standard scores for each teacher. Since the frequency distributions of these four measures for the total sample were approximately normal, it was not necessary to normalize. A mean of the four z-scores for the selector measures was computed for each teacher. An attempt was made to select the two highest controlling teachers and the two lowest controlling teachers within each grade based upon the mean z-score for the four selector measures; however, a number of teachers had moved out of the state or to other school districts since the completion of the earlier study and this was not possible in all grades. In cases where the most extreme teachers were no longer available for study, the next most extreme teachers were selected. One extra high control teacher was selected in grade 5 as an alternate for some analyses where data were missing on the original teacher. The z-scores on the selector measures for the nine high controlling teachers and the eight low controlling teachers were .85 and -.78 respectively. The mean z-scores of the high controlling teachers ranged from .53 to 1.52, and the mean z-scores of the eight low controlling teachers ranged from -.29 to -1.19. As these scores indicate, there was no overlap between the two groups of high and low controlling teachers, and these groups represented the extremes on the measures of control used.

In addition to the main sample of 17 teachers, two teachers from the original sample were selected who exhibited much classroom controlling behavior, but who also exhibited much warmth and affection towards their pupils. Positive correlations were obtained in the earlier investigation between measures of teacher classroom controlling behavior and measures of the teachers' aloofness towards pupils; however, there were several teachers who appeared to deviate from these relationships. Two teachers were selected who scored high on the Teacher Preference Schedule measure of control and high on the ratings and verbal statements measures of controlling behavior, but who scores at the low end of the factor score defined as warm, permissive vs. cold, controlling. It was hypothesized that such a profile of control scores might describe a teacher who showed much warmth and affection towards her pupils in addition to high controlling behavior. The creativity gain scores of the pupils of these two teachers were compared to the gains achieved by the pupils of the highly controlling teachers in the main sample.

Pupil Measures

Creativity pre-test and post-test, and the measure of intelligence. All pupils of the 19 teachers described above were tested during the 1962-1963 academic school year. During the last week in November, the pre-test of the Torrance battery of creativity tests was administered. In administering these tests, the group administration procedure was used. Only the pupils in grades four and five received the verbal battery of creativity tests, since this battery requires pupils to write out their responses. All pupils in grades two through five received the non-verbal creativity battery.

The post-test creativity measures were administered during the third week in May, 1963. The post-test creativity measures consisted of equivalent forms of the pre-test batteries. The fourth and fifth graders again received the verbal battery, while all grades received the non-verbal battery.

The creativity tests were scored according to the scoring procedures described in the manual provided for these tests by Yamamoto (1962). Interscorer reliability coefficients were computed for two independent scorers. The interscorer reliabilities compared favorably to those reported by Torrance for these creativity tasks. The interscorer reliability coefficients are reported in the following chapter.

Correlations were computed between the creativity pre-test and post-test scores to provide data on the equivalent form, test-re-test reliability of these tests. Reliability coefficients were computed separately at each grade level, for subtest and battery total scores. The data provided comparisons of the reliabilities at the different grade levels two through five and comparisons of the reliabilities of the different subtest scores and battery total scores.

The reliability coefficients, and other relationships reported in the study were computed on the IBM 7094 high speed electronic computer at Western Data Processing Center, The University of California at Los Angeles. A missing data correlational program was used in the present study, necessitated by the fact that complete data was not available for all of the pupils in the sample. For this reason the sample sizes varied slightly for correlations between the different variables in the study. The number of students for which both the creativity pre-test and creativity post-test scores were available ranged from 158-183 in grade 2, 111-113 in grade 3, 105-120 in grade 4, and 132-147 in grade 5.

During the month of January 1963 all pupils in the sample were administered the Lorge-Thorndike group intelligence test, Form A. The Lorge-Thorndike non-verbal battery was administered to all pupils in grades two through five. The Lorge-Thorndike verbal battery was also administered in grades four and five. Correlations between the Lorge-Thorndike intelligence scores and scores on the creativity pre-test provided data on the independence of the creativity measure and a measure of verbal and non-verbal intelligence. The Lorge-Thorndike measure was also used as a supplementary variable in subsequent analyses to control the factor of intelligence in testing hypotheses concerning the creativity measures.

Creativity Gain Scores. To obtain creativity gain scores, intercorrelations between scores on the creativity pre-test, the intelligence measure, and the creativity post-test were computed within each of the four grade levels studied. The predicted creativity post-test scores were computed from a multiple regression equation and then subtracted from the obtained post-test score, providing the adjusted creativity gain scores. The predictors in the multiple regression equation were the creativity pre-test scores and intelligence, and the criteria were the creativity post-test scores. The adjusted gain scores were computed separately within each of the four grades studied. In grades two through five, non-verbal creativity gain scores were computed for the Fluency, Flexibility, Originality, Elaboration, and the Total non-verbal scores. In grades four and five, adjusted gain scores were computed for the verbal creativity battery for Fluency, Flexibility, Originality, Elaboration, and Total verbal scores. A high adjusted gain score for a given pupil indicated that the pupil achieved a higher creativity post-test score than was predicted for him on the basis of his performance on the creativity pre-test and intelligence measures.

The adjusted creativity gain scores served as the dependent variables in one-way classification analyses of variance which were computed for each creativity test score within each grade level studied. The main effect of interest in these analyses was the between high and low controlling teachers main effect. The test of this effect provided an indication of the extent to which teacher controlling behavior in the classroom affects pupil gains in creativity test performance over the school year. A simple one-way classification analysis of variance design was employed in this phase of the study rather than a more complex design because of the fact that the number of pupils in the classrooms tested differed. Unequal numbers of subjects can be easily treated in a one-way classification design since the analysis involves no interactions.

Teacher controlling behavior, pupil creativity, and pupil classroom behavior. To investigate the effects of teacher classroom controlling behavior, pupil creativity, and their interaction effects on pupil behavior in the classroom, two high creative and two low creative pupils were selected on the basis of their scores on the Torrance battery of creativity tests for each teacher in grades four and five. The subjects employed in this analysis were matched for intelligence. It was anticipated that there might be sex differences associated with the behavior of pupils in the classroom. Hence, it was decided to observe boys only.

In an extensive review of classroom observational procedures, Medley and Mitzel (1963) reported that the reliability of such measures of classroom behavior are frequently very low when only a few visits are made to the classroom. They found that reliability may increase substantially when the number of visits are increased. Each observer obtained three 15-minute observational sessions per pupil bringing the total number of observational sessions to six per pupil.

As mentioned above, two high and two low creative boys were selected on the basis of their creativity test scores in each of eight fourth and fifth grade classrooms taught by the teachers in the sample. The pupils to be observed were selected by a third party so that the observers had no knowledge of the group in which a pupil fell. Since the teachers in grades four and five had already been dichotomized in terms of their classroom controlling behavior, a 2 x 2 x 2 factorial design was employed. High and low teacher controlling behavior constituted one factor, high and low pupil creativity as measured by the Torrance battery constituted the second factor, and grade level (grades four and five) the third factor. A total sample of 32 pupils were observed individually.

The dependent variables in this analysis were 12 dimensions of observed pupil classroom behavior. These 12 dimensions of pupil behavior represented a classification of 39 items of pupil behavior which were recorded by the two independent observers. To control the effect of specific classroom activities on pupil behavior, the order in which the pupils were observed was randomized. Each 15-minute observation period was further divided into five three-minute periods. During each three-minute period, the observer checked on the observational record sheet all the behaviors that he observed which were emitted by the pupil. The pupil observation record with the complete list of pupil behavior items is reproduced in Appendix C. The 12 dimensions of pupil behavior which were derived from the 39 items were as follows:

Pupil Questioning Behavior - Q
Pupil Academic Behavior - A
Pupil Management Behavior - M
Pupil Leadership Behavior - L
Pupil Non-conformity Behavior - NC
Pupil Talk - PT
Pupil Grouping - G
Pupil/Teacher Interaction - I
Pupil Emotional Behavior - E

Pupil Ideas - Id
Self-initiated Pupil Academic Behavior - SA

Self-initiated Pupil Non-academic Behavior - SNA

This list was compiled from a number of instruments designed to measure classroom behavior. In most cases the instruments reviewed had been designed to measure the classroom behavior of the teacher, but in several cases pupil behavior was also assessed. Those devices which provided for the assessment of pupil classroom behavior usually assessed the behavior of the class as a whole, but many of the behavioral items were easily translated into a form appropriate for observing individual pupils. The behavioral items in the observation schedule developed for the present investigation were adopted from The Flanders Interaction Matrix (Flanders, 1962), the Observation Schedule and Record or OSCAR (Medley and Mitzel, 1958), and the Observational Record used by Ryans (1960) in his study of teacher characteristics. In addition to these sources, the authors included several student behavior dimensions of their own.

Scores were derived for the above dimensions by totaling the number of behavioral items checked under each dimension for the six 15-minute observation periods.

In addition to actually recording observed pupil behavior, the observers also rated each pupil on four rating scales after each 15-minute observational session. Seven-point rating scales were used for the four dimensions: Apathetic - Alert, Obstructive - Responsible, Uncertain - Confident, and Dependent - Initiating. Definitions of these four dimensions in behavioral terms are also reported in Appendix C.

After the observers had completed their observations of the pupils, each teacher was asked to rate the four pupils in her class on the four rating scales described above. It was anticipated that the teachers' ratings and the observers' ratings would provide some interesting comparisons. The teachers also supplied the observers with information concerning the number of days each pupil had been absent during the school year up to the time the observations were completed.

The complete observational procedure was pre-tested on a number of pupils in grades four and five at the laboratory school on the University of Utah campus. During the pre-test observations, both observers independently observed the same pupils at the same time. This provided information concerning the observer agreement achieved in employing the observational record of pupil behavior. The 32 pupils who were selected on the basis of high and low scores on the Torrance creativity battery were then observed during the month of April, 1963. Data on the reliability or stability of the dimensions of classroom behavior were obtained by correlating the scores derived from the three observations of one observer with the scores derived from the three observations of the other observer. The resulting reliability coefficient reflected both errors in observer agreement and the instability of pupil behavior over time.

Since the order of observing pupils was randomized, the time interval between observation sessions varied for different pupils. The time interval between observation sessions ranged from one day to one week.

CHAPTER VI

PHASE II: RESULTS

The Effect of Teacher Controlling Behavior in the Classroom On Pupil Gains in Creativity as Measured by the Torrance Tests

Pupil creativity gain scores adjusted for the creativity pretest and intelligence were computed for each pupil. The pupil gain scores were computed on the basis of within grade regression equations. Adjusted pupil gain scores were computed for the following ten scores derived from the Torrance creativity battery: Non-verbal Fluency, Flexibility, Originality, Elaboration, and Total, and Verbal Fluency, Flexibility, Originality, Elaboration, and Total scores. These ten adjusted gain scores were used as the dependent variables in an analysis of variance design. One-way classification analyses of variance were computed for each non-verbal score within each grade two through five to compare the creativity test gains of high and low controlling teachers. Analyses of variance of the verbal creativity gain scores were computed within grades four and five.

The means and standard deviations of the Total Non-verbal Creativity gains and the Total Verbal Creativity gains are shown in Table 6.01 for each of the 19 teachers studied. A constant of 200.5 was included in the formula for computing the adjusted gains to eliminate negative gain scores. A mean gain of 200.5 indicates that on the average, the pupils of a given teacher achieved creativity post-test scores which were identical to the post-test scores predicted for them on the basis of their scores on the creativity pre-test and the intelligence measure. The means shown in Table 6.01 indicate that the pupils of some teachers achieved creativity post-test scores above those predicted for them, while the pupils of other teachers achieved post-test scores below those predicted for them.

One of the assumptions underlying the analysis of variance used in the present study is that of homogeneity of variance within the cells of the design. To test this assumption in the present data, the variances of the pupil gains for the classes within each grade were compared by means of Hartley's maximum F-test for homogeneity of variance. The hypothesis of homogeneity of variance was not rejected for the Total Non-verbal Creativity gain scores in any of the grades. A comparison of the variances of the Total Verbal Creativity gains produced an F significant at the .05 level in grade four and an F significant at the .01 level in grade five. The ratio of the largest variance to the smallest variance in grade four was approximately three to one. A study by Boneau (1960) indicated that violations of the assumption of homogeneity of variance of the magnitude four to one did not seriously bias the t-test when the sample sizes were equal and fairly large (e.g., Boneau's N's were fifteen per treatment group). Since the t-test is directly analogous to the F-test in a simple analysis of variance, Boneau's results are

TABLE 6.01

MEANS AND STANDARD DEVIATIONS OF THE REGRESSED CREATIVITY TEST GAIN SCORES FOR EACH TEACHER*

TEACHER NUMBER	TOTAL NON-VERB. CREAT.			TOTAL VERB. CREAT.	
	N	$\overline{X}$	s	$\overline{X}$	s
Grade 2:					
High Control					
1	19	216.80	24.18		
2	27	189.77	20.67		
Low Control					
3	28	205.19	15.70		
4	19	193.09	17.94		
Warm/High Control					
5	24	210.39	27.04		
6	27	190.93	22.56		
Grade 3:					
High Control					
7	24	197.09	22.34		
8	29	200.25	23.85		
Low Control					
9	26	220.80	27.45		
10	26	182.89	19.04		
Grade 4:					
High Control					
11	25	203.45	26.22	206.09	37.24
12	22	200.48	18.76	178.81	21.80
Low Control					
13	27	197.57	16.22	209.62	26.13
14	25	198.93	15.92	201.84	27.53
Grade 5:					
High Control					
15	21	191.83	23.27	199.74	24.26
16	30	209.63	25.49	194.69	35.08
17	19	203.91	24.91	208.09	52.10
Low Control					
18	21	200.33	23.43	197.72	45.07
19	31	193.39	21.56	195.05	24.78

*The gain scores include a constant of 200.5 to eliminate negative numbers.

quoted here to defend the position that the heterogeneity of variance of the magnitude three to one obtained for the Total Verbal Creativity gains in the fourth grade did not lead to spurious results. The number of pupils in the four fourth grade classes were nearly equal ranging from 22 to 27, and the N's were somewhat larger than those employed in the Boneau study. In the fifth grade, the ratio of the variances was larger as were the discrepancies in sample sizes suggesting a greater possibility of bias in the F-test. As a precautionary measure, the analyses of the Verbal Creativity gains in the fourth and fifth grades were repated using equal sample sizes. Subjects were eliminated by means of a table of random numbers to provide equal N's per cell. The equal N's analyses produced results similar to the results obtained in the initial analyses.

In addition to the tests of the assumption of homogeneity of variance discussed above, the frequency distributions of the creativity gain scores of the different classes were alos examined. Although the N's of the class distributions were often quite small, in general the distributions appeared to be approximately normal in shape. In one or two exceptions, a slight positive skew appeared in the distributions; however, these exceptions were not considered to be large enough to affect the results of the analyses of variance.

Thirty analyses of variance were computed comparing the pupil gains in the various creativity test scores for high and low controlling teachers. Two high and two low controlling teachers were compared within each grade with the exception of the grade five where data was available for three high controlling teachers and two low controlling teachers.

The analyses of the Verbal Creativity gains in grade five produced no significant teacher control main effects. The analysis of the Verbal Creativity gains in the fourth grade, however, produced several statistically significant F's for the main effect of teacher controlling behavior in the classroom. Table 6.02 contains the summary of the analysis of variance of the Total Verbal Creativity gains in grade four. An F of 4.67 significant at the .05 level of confidence was obtained for the teacher control main effect. The direction of the difference indicated that the pupils of teachers who exhibited little classroom controlling behavior achieved higher gains in verbal creativity than the pupils of teachers who exhibited much classroom control. An examination of the mean gains indicated that the pupils of the highly controlling teachers performed on the average lower than their predicted post-test scores, whereas the pupils of the low controlling teachers performed higher than their predicted scores. This analysis was repeated with equal numbers of pupils per cell and an F of 9.53 was obtained for the teacher control main effect which was significant at the .01 level of confidence. The direction of the effect was the same as that obtained in the first analysis. The larger F of the second analysis resulted from the fact that the random elimination of subjects to equalize the N's produced a chance increase in the difference between the means of the treatment groups.

Table 6.03 summarizes the analysis of variance of the Verbal

TABLE 6.02

ANALYSIS OF VARIANCE OF TOTAL VERBAL CREATIVITY GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE FOUR

SOURCE	d.f.	F	P
Between Teachers	3	5.35	.01
Between Control	1	4.67	.05
Residual Between Teachers	2	5.69	.01
Within	95	(833.75)[a]	
Total	98		

[a] Equals the mean square of the error term.

Flexibility gains in the fourth grade. This analysis produced a highly significant F at the .005 level of confidence for the teacher control main effect. The pupils of the low controlling teachers again achieved higher Verbal Flexibility gains than the pupils of the highly controlling teachers.

Finally, Table 6.04 summarizes the results of the analysis of variance of the Verbal Fluency gains in grade four. The F for the teacher control main effect in this analysis was significant at the .10 level of confidence. The pupils of low controlling teachers again achieved the higher gains in Verbal Fluency.

The results of these analyses indicated that the pupils of low controlling teachers in the fourth grade achieved greater gains on the Torrance measure of verbal creativity than did pupils of highly controlling teachers. The effect of teacher classroom control was particularly marked in the case of the Verbal Flexibility score derived from the tests. The Verbal Flexibility score may best be described as a measure of a pupil's flexibility in the flow of ideas produced on the test. For example, on the Unusual Uses test the pupil is asked to write down as many unusual uses he can think of for a toy dog other than as a plaything. Each response is classified into one of three categories; namely, Ornamental-Household, Container-Holder, and Scientific-Mechanical. The lowest possible Flexibility score would be obtained when a pupil's responses all fell into one category. The Flexibility score is obtained by counting the number of runs of the three categories of responses throughout the series of responses produced by the pupil. A high Flexibility score would be assigned when a pupil frequently used all three categories of response, and when his responses frequently jumped around from category to category. The finding that the pupils of highly controlling teachers achieved lower gains in flexibility than the pupils of low controlling teachers was consistent with what would be expected from classrooms taught by highly controlling teachers. Such teachers typically prefer a highly ordered and controlled classroom. Variability and flexibility in pupil behavior may pose a threat to the teacher's attempt to maintain order and control in the classroom, and the teacher may attempt to elim nate such behavior by means of reprimands and disapproval. The highly controlling teacher may positively reinforce fonformity on the part of the pupils, and pupil behavior which is consistent with the teacher's desire for classroom order and control.

The absence of a statistically significant effect of teacher controlling behavior on pupil gains in verbal creativity in the fifth grade may have resulted from the fact that the fifth grade teachers in the present sample did not have sufficient opportunity to influence their pupils. Three out of five of the fifth grade teachers taught their home room classes only an hour-and-a-half per day. Pupils in these classes frequently traveled to other classrooms for training in English and mathematics and came into contact with many more of the school personnel than did the pupils in the lower grades. It is possible that the influence of the home room teacher with regard to pupil gains in verbal creativity was offset by other influences in the school. It is also possible that pupils by the time they reach the fifth grade are

TABLE 6.03

ANALYSIS OF VARIANCE OF VERBAL FLEXIBILITY GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE FOUR

SOURCE	d.f.	F	P
Between Teachers	3	6.18	.001
Between Control	1	11.43	.005
Residual Between Teachers	2	3.55	.05
Within	95	(33.52)[a]	
Total	98		

[a] Equals the mean square of the error term.

TABLE 6.04

ANALYSIS OF VARIANCE OF VERBAL FLUENCY GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE FOUR

SOURCE	d.f.	F	P
Between Teachers	3	5.88	.01
Between Control	1	3.84	.10
Residual Between Teachers	2	6.90	.005
Within	95	(111.96)[a]	
Total	98		

[a] Equals the mean square of the error term.

less amenable to teacher influence than pupils in the lower grades. The results of the analyses of variance within the different grades tended to support the conclusion that the fifth grade teachers had little influence on the verbal creativity gains of their pupils. With the exception of the fifth grade verbal creativity analyses, practically all other analyses produced statistically significant between teachers or residual between teachers' F-tests even when the high and low teacher control effect was nonsighificant. Such a finding indicates that some unknown differences among the teachers or pupils produced differences in creativity test gains. In the fifth grade, however, no differences between teachers were found for pupil gains in verbal creativity, suggesting that no teacher influence of any kind occurred at least with regard to gains in verbal creativity.

The above argument concerning the lack of teacher influence on pupil gains in verbal creativity in the fifth grade is weakened to some extent by the finding of one significant difference obtained for a score derived from the non-verbal creativity battery in the fifth grade. The analyses of variance of the non-verbal creativity gains within the four grades produced one statistically significant main effect for high and low teacher controlling behavior. A significant difference between high and low controlling teacherswas obtained for pupil gains in the Non-Verbal Elaboration score in grade five. The summary of this analysis is shown in Table 6.05. When the direction of the difference was examined in this analysis, it was found that the pupils of high controlling teachers actually achieved the greater gains on the Non-Verbal Elaboration score than did the pupils of the low controlling teachers. At first glance, this reversal in the effect of teacher controlling behavior would appear to conflict with the finding that low controlling teachers achieved the higher gains in verbal creativity; however, the reversal is the teacher control effect for the Non-Verbal Elaboration score is not as surprising when the nature of the Elaboration score is taken into account. The non-verbal creativity battery requires a pupil to draw pictures incorporating certain stimulus figures into the drawings. A high Elaboration score would be assigned to a drawing when the drawing contained much detail. For example, a simple outline of a house (e.g., a square with a triangle for a roof) would receive a low Elaboration score, whereas a picture of a house containing windowns, doors, chimney, shingled roof, landscaping, etc. would receive a high Elaboration score. In accounting for the present results, it is possible that the highly controlling teacher pays greater attention to the details and accuracy of pupil products, encouraging more elaborate pupil responses, while the low controlling teacher pays greater attention to the nature and diversity of the pupils' ideas.

It will be recalled that two second grade teachers were included in the study whose scores on the measures of teacher classroom control appeared to describe a teacher who showed much warmth towards her pupils in addition to much controlling behavior. The non-verbal creativity test gains of the pupils of these two teachers were compared to the gains of pupils in the classes of teachers who were classified as highly controlling, cold, and aloof. Analyses of these data produced no significant

TABLE 6.05

ANALYSIS OF VARIANCE OF NON-VERBAL ELABORATION GAINS FOR HIGH AND LOW CONTROLLING TEACHERS IN GRADE FIVE

SOURCE	d.f.	F	P
Between Teachers	4	3.89	.01
Between Control	1	4.28	.05
Residual Between Teachers	3	3.77	.05
Within	117	(55.45)[a]	
Total	121		

[a] Equals the mean square of the error term.

differences in the gains of the two groups. The warm but controlling teachers were also compared to the low controlling teachers in the second grade and again no significant differences were obtained.

In addition to the statistical control of the pupils' level of performance on the creativity pre-test and intelligence measures provided by the use of adjusted gain scores, the mean pre-test and intelligence scores of the high and low teacher control treatment groups were examined. The comparisons indicated whether the pupils of high and low controlling teachers started at approximately the same levels on the creativity tests which were administered at the beginning of the school year. The means were nearly identical in all comparisons. The mean raw score on the intelligence measure was 114.39 for the low control classes in grades four and five as compared to 112.50 for the high control classes in these grades. The mean raw score on the intelligence measure was 50.66 for the low control classes in grades two and three as compared to 50.25 for the high control classes in these grades. The mean verbal creativity pre-test was 80.99 for the low control groups, and 85.63 for the high control groups. A t-test of this difference produced a t of less than one. The non-verbal creativity pre-test mean for the low control classes was 110.57 as compared to 110.92 for the high control classes. As a result of the statistical control of the pupils' performance on the creativity pre-test and intelligence measures, and in light of the similarity in the above means, it can be concluded the differences obtained in the pupil creativity gains of high and low controlling teachers were not due to differences in the pupils' initial level of performance on the creativity tests or to differences in pupil intelligence.

In conclusion, the results of the analysis of pupil gains on the Torrance measures of creative thinking indicated that low controlling teachers in the fourth grade achieved greater gains in verbal creativity and flexibility than did the high controlling teachers. On the other hand, highly controlling teachers in the fifth grade tended to achieve greater pupil gains on a measure of elaboration derived from the drawings of pupils on the non-verbal creativity battery.

Relationships Between Pupil Classroom Behavior, Pupil Creativity, and Teacher Controlling Behavior

As mentioned in the previous chapter, the observation schedule used to record pupil classroom behavior was tried out on a small sample of fourth and fifth grade pupils at the laboratory school on the University of Utah campus. The try-out of the observational schedule provided data on observer agreement, enalbed the observers to familiarize themselves with the record and the observational procedure, and indicated whether the schedule provided a workable procedure. Twenty pupils were observed individually by two independent observers. The observers observed the same pupil at the same time so that information concerning observer agreement could be obtained. Each pupil was observed for one 15-minute session. An examination of the tallies in the 39 behavior categories made by the observers indicated that 74 per cent perfect agreement was achieved. Since three days were required to obtain the data on the try-out sample of 20 pupils, an analysis of the percentage of perfect agreement for each separate day indicated that the observers tended to improve in their use of the observational schedule. On the first day, the observers achieved 71 per cent perfect agreement, on the second day 73 per cent, and on the third day 78 per cent perfect agreement.

The scores for the major behavioral categories were obtained by counting the number of tallies of specific behaviors which were include in each main category. For a number of categories, this score was zero since no behaviors included in the categories were observed. The lack of behaviors recorded in certain categories may have been due to the short period of observation, or to the rarity of certain pupil behavior. Five of the main categories of pupil behavior contained sufficient tallies to provide meaningful correlations between the scores derived from the records of the two independent observers. Coefficients of observer agreement were obtained for these five categories as follows: Academic Behavior .82, Non-conformity Behavior .87, Pupil Talk .56, Pupil Works Alone .89, and Pupil Works in a Group .91. The coefficients of inter-rater agreement for the four rating scales were as follows: Obstructive-Responsible .75, Apathetic-Alert .52, Uncertain-Confident .23, and Dependent-Initiating .51. In spite of the small sample of trial data and the short observation period, the data indicated that the observers' records of pupil behavior showed considerable agreement. The agreement obtained for the rating scales was considerably less than that for the behavioral records, a finding which was consistent with the findings of other studies employing such techniques.

The reader will recall that after the initial try-out of the pupil observation schedule, pupils were selected from the main sample of subjects who scored high and low on the measures of creative thinking, but who were matched on a measure of intelligence. A total of 32 pupils were selected, and each observer observed each pupil for three 15-minute observation sessions. Data on the reliability of the dimensions of pupil behavior was provided by correlations between the scores derived from the three 15-minute observation periods of observer one and the scores derived from the three observation periods of observer two. For the present purposes the term reliability coefficient refers to correlations between pupil behavior scores obtained by different observers at different times, as distinguished from coefficient of observer agreement which reflects behavior observed by different observers at the same time. Little or no behavior was observed in a number of categories, and no estimate of reliability could be obtained for these categories. Categories for which insufficient behavior was observed were Questioning, Leadership, Emotional Behavior, Ideas, Self-initiated Academic Activity, and Self-initiated Non-academic Activity. The reliabilities of the categories for which sufficient behavior was observed are shown in Table 6.06, in addition to the reliabilities of the four observer ratings. Since the pupil behavior scores used in the later analyses were the combined totals of the two observers' scores, the Spearman-Brown formula was employed to estimate the reliability of the total scores. The

TABLE 6.06

THE RELIABILITY OF NINE CATEGORIES OF PUPIL BEHAVIOR AND THE RATINGS OF PUPIL BEHAVIOR

N = 32	R (OBSERVER 1 VS. OBSERVER 2 TOTALS)	R (ESTIMATED FOR TOTAL SCORE: SPEARMAN-BROWN)
Academic	-.14	
Non-conformity	.70	.82
Pupil Talk (Self-Initiated)	.39	.56
Pupil Talk (Responds to Teacher)	.06	
Pupil Works Alone	-.10	
Pupil Works in Group	.00	
Teacher-Pupil Interaction (Initiated by Pupil)	.35	.52
Teacher Pupil Interaction (Initiated by Teacher)	-.16	
Management	.09	.17
Ratings:		
Apathetic-Alert	.32	.48
Obstructive-Responsible	.51	.68
Uncertain-Confident	.20	.33
Dependent-Initiating	.13	.23

reliabilities shown in column one of Table 6.06 are analogous to correlations between two halves of a test, while those in column two are analogous to the estimated reliability for the score derived from the whole test. The use of the Spearman-Brown estimate in the present case of pupil behavior scores rests on the assumption that the errors in the scores derived from the observations of each observer did not differ. Although there was no way to check this assumption in the data available, there was no reason to expect the errors of the two observers to differ.

The reliabilities in Table 6.06 indicated that only three of the pupil behavior dimensions were sufficiently reliable, and that the Non-conformity dimension was by far the most reliable of the three. The Non-conformity dimension may be described as a dimension of "acting up" behavior on the part of the pupil. The results suggest that acting up or misbehavior is a fairly stable dimension of pupil classroom behavior and is relatively easy to record by classroom observers. The other two somewhat less reliable behavior dimensions were Self-initiated Pupil Talk, and Teacher Pupil Interaction Initiated by the Pupil. Three of the four observer ratings evidenced some reliability: Apathetic-Alert, Obstructive-Responsible, and Uncertain-Confident. The Obstructive-Responsible rating was the most reliable of the ratings tending to support the notion that pupil misbehavior is relatively stable from day to day and most easily assessed by classroom observers.

Although the reliabilities of these pupil classroom behavior measures were low when compared to the reliabilities of more traditional psychological tests, they compared favorably to reliabilities obtained from other measures of classroom behavior (see Medley and Mitzel, 1963). Rarely do measures of classroom behavior provide reliabilities above .50, and when higher reliabilities are achieved, they are usually a result of increased numbers of visits to the classroom or increased numbers of observers.

It will be recalled that teacher ratings were obtained on the four rating scales for the 32 pupils in the study of pupil classroom behavior. Correlations between teacher ratings and observer ratings indicated little relationship between pupil behavior rated by observers and pupil behavior rated by teachers. These correlations for the four rating scales were .09 for Apathetic-Alert, .20 for Obstructive-Responsible, .14 for Uncertain-Confident, and .31 for Dependent-Initiating. None of these correlations were significant at the .05 level of confidence; however, one (Dependent-Initiating) was significant at the .10 level. The results indicated that classroom observers and teachers did not agree on ratings of pupil classroom behavior. The fact that the observers' ratings evidenced some reliability suggested that the teachers' ratings were either unreliable or that the teachers were responding to pupil behaviors other than those listed in the glossary of behaviors which accompanied the rating scales. Several of the teachers' ratings also failed to correlate with the dimensions of recorded pupil behavior The Obstructive-Responsible teacher rating correlated .06 with pupil Non-conformity behavior, and the Dependent-Initiating teacher rating correlated -.03 with Self-initiated Pupil Talk.

Further analysis of the pupil behavior scores and ratings was made only for scores which had demonstrated sufficient reliability. The scores employed in further analyses and as the dependent variables in the analyses of variance to be reported were: Non-conformity, Self-initiated Pupil Talk, Teacher-Pupil Interaction Initiated by Pupil, observer ratings for Apathetic-Alert, Obstructive-Responsible, Uncertain-Confident, and Dependent-Initiating, and the four teacher ratings. (Since there was no way to compute the reliabilities of the teacher ratings, they were included in the analyses.) One additional score, the number of days the pupils were absent throughout the year, was also included.

Table 6.07 contains the correlations between intelligence and creativity test total scores and the pupil behavior measures described above. As can be seen in the table, neither intelligence nor creativity correlated significantly with the three reliable dimensions of recorded pupil behavior. Consistently low positive correlations were obtained between intelligence and the ratings of pupil behavior. The correlations between intelligence and teacher ratings were somewhat higher and more significant than the correlations between intelligence and observer ratings. This finding suggests that the teachers may have given greater weight to pupil intelligence or to correlates of intelligence in making their ratings than did the observers. The correlations between the measures of pupil behavior and creativity were, with one exception, nonsignificant. The exception was the correlation between creativity and the Apathetic-Alert observer rating which was .36 and significant at the .05 level.

Twelve analyses of variance were computed using each of the twelve measures listed at the left hand side of Table 6.07 as a dependent variable. It will be recalled from the previous chapter that the design employed to analyze the pupil behavior measures was a 2 x 2 x 2 factorial design. Factor 1 consisted of two groups of high and low creative pupils matched for intelligence; factor 2 consisted of high and low teacher classroom control, and factor 3 consisted of grade level (e.g., grades four and five). The analyses of variance of nine of the measures of pupil classroom behavior produced non-significant results. The nine measures for which nonsignificant results were obtained were: Teacher/Pupil Interaction Initiated by the Pupil, observer ratings for Dependent-Initiating, Uncertain-Confident, and Obstructive-Responsible pupil behavior, teacher ratings for Apathetic-Alert, Uncertain-Confident, Obstructive-Responsible and Initiating pupil behavior, and the number of days absent. Analysis of the remaining three measures of pupil classroom behavior produced several F's approaching significance which were merely of suggestive interest.

The analysis of variance of the measure of pupil non-conformity behavior is shown in Table 6.08. The only significant F in this analysis was obtained for the teacher classroom control by grade level interaction. This finding was of secondary interest since it was not related to any of the major hypotheses of the study.

TABLE 6.07

RELATIONSHIPS BETWEEN PUPIL CLASSROOM BEHAVIOR AND MEASURES OF PUPIL INTELLIGENCE AND CREATIVITY

N = 32	INTELLIGENCE	TOTAL CREATIVITY
Non-Conformity	-.21	.03
Self-Initiated Pupil Talk	-.01	-.20
Teacher/Pupil Interaction Initiated by the Pupil	-.14	-.09
Observer Ratings:		
Alert	.26	.36*
Responsible	.20	.05
Confident	.19	.27
Initiating	.36*	.11
Teacher Ratings:		
Alert	.41*	-.05
Responsible	.24	-.12
Confident	.44*	-.05
Initiating	.36*	.06
Number of Days Absent	-.18	.27

*Significant at the .05 level (two-tailed test).

TABLE 6.08

ANALYSIS OF VARIANCE OF THE NON-CONFORMITY BEHAVIOR IN THE CLASSROOM FOR HIGH AND LOW CREATIVE PUPILS OF HIGH AND LOW CONTROLLING TEACHERS

N = 32

SOURCE	d.f.	F	QUASI-F
Between Creative Groups	1	.12	
Between Teachers	7	5.71	
Between Control	1	.03	
Between Grade	1	.05	
Control X Grade	1	6.97*	7.22*
Residual Between Teachers	4		
Interactions:			
Creativity X Teacher	7	.70	
Creativity X Control	1	.65	
Creativity X Grade	1	.74	
Creativity X Control X Grade	1	.08	
Residual Between Cells	4	.86	
Within (Error Term)	16	(37.94)[a]	
Total	31		

*P $<$.025

[a] Equals the mean square of the error term.

Some comment should be made concerning the nature of the entries in the summary of the analysis of variance shown in Table 6.08. For ease of presentation and reproduction, only the sources of variance, degrees of freedom and F-ratios are shown. The mean square for the error term is shown in parentheses so that by simple multiplication the entire table may be reconstructed. The mean squares for each component may be obtained by multiplying each F by the error mean square, and the sums of squares may be obtained by multiplying each mean square by its corresponding degrees of freedom.

Table 6.08 also contains a column headed by the term quasi-F. This column contains F-ratios based upon the error term adjusted for the correlation between intelligence and the dependent variable. Although the subjects selected on the basis of high and low creativity test scores were matched on intelligence, intelligence could still correlate with the dependent variables within cells (see Table 6.07). It was therefore possible to reduce the error term taking out the covariance between the dependent variable and intelligence. To save on the labor of computing twelve analyses of covariance the quasi-F procedure was used as follows:

$$\text{Regular F} = \frac{\sigma_e^2 + n\,\sigma_t^2}{\sigma_e^2}$$

$$\text{Quasi-F} = \frac{\sigma_e^2(1 - r^2) + n\,\sigma_t^2}{\sigma_e^2(1 - r^2)}$$

The adjusted error term is equal to $\sigma_e^2(1-r^2)$ where r is the correlation between intelligence and the dependent variable. This formula gives the amount of error variance that remains after the percentage of the error variance predictable from the intelligence measure was taken out. The adjusted error variance was then substituted for the unadjusted error term in both the numberator and denominator of the F-ratio. In considering the results which follow, the reader should keep in mind that the quasi-F procedure used in the present study was biased in that it produces a greater number of significant F's than would be obtained from the usual analysis of covariance procedure. The main reason for the bias is that in analysis of covariance one degree of freedom is lost in the error term due to the restriction of the regression line, and this loss of one degree of freedom was not taken into account in making the adjustments for the quasi-F test in the present case. The experimenters were willing to live with the bias of the quasi-F test in the light of the tremendous amount of time saved by the procedure. Since the correlations between intelligence and the dependent variables were generally quite small, the quasi-F's were only slightly larger than the original F's, and no great reductions occurred in the probability values. If the correlations between the dependent variables and intelligence had been larger, the more usual analysis of cov ariance procedure would have been employed. (For a general discussion of quasi-F ratios see Winer, 1962.)

The significant teacher control by grade level interaction reported in Table 6.08, although difficult to interpret, suggests that the effect of teacher classroom control on pupil non-conformity or misbehavior depends upon the grade level of the pupils. Specifically, the pupils of more permissive teachers exhibited less misbehavior in the fourth grade than in the fifth grade, whereas just the reverse occurred for the highly controlling teacher. Pupils of highly controlling teachers exhibited much misbehavior in the fourth grade and a sharp drop in misbehavior in the fifth grade. One interpretation of these results is that permissive teachers are more effective in eliminating misbehavior in the lower grades, while highly controlling teachers are more effective in eliminating misbehavior in the upper grades. This interpretation must be considered tentative since only two grade levels were examined in the present study. For a more conclusive interpretation, the effects of teacher controlling behavior on pupil behavior must be examined across the entire range of school grades.

The results reported in Table 6.08 failed to support the hypothesis that high and low creative pupils as distinguished by the Torrance tests would differ on a measure of pupil non-conformity behavior in the classroom. In addition, a significant pupil creativity by teacher control interaction was not obtained. Nonsignificant findings resulted from this analysis in spite of the fact that the pupil non-conformity dimension of classroom behavior was the most reliable pupil behavior measure (e.g., R = .82).

The second analysis of variance to be reported is that for the Self-initiated Pupil Talk dimension of pupil classroom behavior. Table 6.09 contains the summary of this analysis of variance. Two F's in this analysis had probability levels at the .10 level of confidence. Although these probability levels were larger than the usual probability criteria for acceptance of hypotheses, they are nevertheless reported as tentative findings. Research conducted in natural classroom setting typically involves high error terms and the experimental variables of such studies must be strong variables in order to produce highly significant effects. In view of the nature of the data of classroom research, the writers believe that probability levels of .10 should not be ignored. As is the case in any research, findings should be replicated and extended, and more refined measuring instruments should be developed to explore the phenomena.

The between high and low creative groups main effect produced an F-ratio with a probability of .10. An examination of the direction of the difference indicated that the high creative pupils as measured by the Torrance battery actually produced less self-initiated talk in the classroom than the low creative pupils. Although this finding tends to support a suppression hypothesis, that is, the self-initiated talk of high creative pupils may be suppressed in the classroom, there was no evidence that teacher controlling behavior affected the self-initiated talk of the low and high creative pupils differentially (e.g., a nonsignificant creativity by control interaction was obtained).

TABLE 6.09

ANALYSIS OF VARIANCE OF THE SELF-INITIATED PUPIL TALK IN THE CLASSROOM FOR HIGH AND LOW CREATIVE PUPILS OF HIGH AND LOW CONTROLLING TEACHERS

N = 32

SOURCE	d.f.		F
Between Creative Groups	1		3.55*
Between Teachers	7		.93
Between Control		1	3.10*
Between Grade		1	.30
Control X Grade		1	.03
Residual Between Teachers		4	.76
Interactions:			
Creativity X Teacher	7		.64
Creativity X Control		1	.18
Creativity X Grade		1	2.31
Creativity X Control X Grade		1	.09
Residual Between Cells		4	.48
Within (Error Term)	16		(8.47)[a]
Total	31		

*P < .10

[a] Equals the mean square of the error term.

The between teacher control main effect produced the second F with a probability of .10 in this analysis, and the finding was more consistent with expectations. An examination of the direction of the difference indicated that pupils of low controlling teachers emitted more self-initiated talk than did the pupils of high controlling teachers. This finding was consistent with the hypothesis that highly controlling teachers are the sole directors of their classrooms and that little self-direction or self-initiated activity occurs on the part of their pupils. On the other hand, the more permissive teacher provides greater opportunity for the pupils to initiate their own activity.

Although the results of the analysis of self-initiated pupil talk were of only borderline significance, the results tie together rather well with the results already reported for the pupil creativity gain score analysis. The pupils of highly controlling teachers tended to emit less self-initiated talk in the classroom and achieved lower gains on measures of verbal creative thinking than did the pupils of low controlling teachers.

Table 6.10 presents the summary of the analysis of variance of the observer rating Apathetic-Alert. Two F-tests were significant beyond the .10 level of confidence in this analysis. The between high and low creative groups main effect indicated that the high creative pupils as measured by the Torrance battery were rated as more alert than the low creative pupils by the classroom observers. In addition, another teacher control by grade level interaction was obtained in this analysis. The pupils of low controlling teachers were rated as more alert in the fourth grade than in the fifth grade. The trend was reversed for pupils of highly controlling teachers with the fourth graders rated as more apathetic, and the fifth graders rated as more alert. This interaction was difficult to interpret in view of the fact that only two grade levels were examined. High controlling teachers may be more successful at eliciting alert pupil behavior in the upper grades as compared to the lower grades. The opposite may be true for more permissive teachers who may be more successful in maintaining alertness of pupils in the lower grades.

The highly tentative nature of conclusions drawn from results having a probability level of .10 must be re-emphasized. The reader should keep in mind that the odds are only 10 to 1 that the above results are due to the treatment effect and not simply due to chance. The results and interpretations already discussed may be most adequately described as results indicating hypotheses for further study and replication.

In summing up the findings of the analyses of the relationships between pupil classroom behavior, measured pupil creativity, and teacher classroom control, it was of interest to note that of 84 F-tests computed, 13 per cent were significant at or beyond the .10 level of confidence; 4.8 per cent were significant at or beyond the .05 level, and 1.2 per cent were significant at or beyond the .025 level. The number of significant F's which were obtained in these analyses was close to the number of "significant" F's which would be expected from a chance distribution

TABLE 6.10

ANALYSIS OF VARIANCE OF THE OBSERVER RATING APATHETIC-ALERT FOR HIGH AND LOW CREATIVE PUPILS OF HIGH AND LOW CONTROLLING TEACHERS

N = 32

SOURCE	d.f.	F	QUASI-F
Between Creative Groups	1	3.27*	3.44*
Between Teachers	7	2.32	
Between Control	1	.34	
Between Grade	1	.00	
Control X Grade	1	4.10*	4.34*
Residual Between Teachers	4	2.96	
Interactions:			
Creativity X Teacher	7	.34	
Creativity X Control	1	.42	
Creativity X Grade	1	1.18	1.19
Creativity X Control X Grade	1	.42	
Residual Between Cells	4	.09	
Within (Error Term)	16	(23.88)[a]	
Total	31		

*P < .10

[a] Equals the mean square of the error term.

of F's. Mention should also be made of the fact that many of the F's which were obtained at the .10 of .05 level resulted from experimental effects which were not central to the main hypotheses of the study. Specifically, several teacher control by grade level interactions were obtained which were difficult to interpret in view of the fact that only two grade levels were examined in this phase of the investigation. The dearth of significant findings related to creativity main effects, teacher control main effects, and creativity by teacher control interactions cannot be written off as entirely the result of unreliable measures. As was mentioned earlier, only the pupil behavior meausres for which some reliability had been demonstrated were employed as dependent variables in these analyses. Evidence has also been presented which indicated that the measures of controlling behavior used to select teachers contained some reliability (e.g., R = .57 for the Aloof-Responsive Q-sort over a one-year period). The creativity Total score used to select the pupils for this design was shown to have a test-retest reliability of .61 in grade four and .75 in grade five. The writer would conclude that the lack of significant results reflects the lack of the validity of the hypothesis that high and low creative pupils as measured by the Torrance battery behaved differently with regard to the dimensions of classroom behavior examined. There was also no evidence that teacher control affects the behavior of high and low creative pupils differentially. The non-significant results for the analysis of the Non-conformity dimension of pupil classroom behavior (e.g., Reliability = .82) provided no support for the common assumption that highly creative elementary school pupils as distinguished by a battery of creativity tests are more non-conforming in the classroom than pupils who score low on such tests. Furthermore, there was no evidence in the present data that pupils who score high on such creativity tests manifest any more or less misbehavior in the class rooms of highly controlling teachers as opposed to the classrooms of more permissive teachers.

The results obtained which had some bearing on the main hypotheses of this phase of the investigation were merely suggestive since the probability values were relatively large (e.g., .10). These results indicated that the highly creative pupils as measured by the Torrance tests emitted less self-initiated talk in the classroom than pupils who scored lower on the tests. When the self-initiated talk of pupils of high and low controlling teachers were compared, the pupils of low controlling teachers tended to emit more self-initiated talk in the classroom than the pupils of highly controlling teachers. Finally, there was a tendency for the high creative pupils to be rated as more alert than the low creative pupils by the classroom observers.

The Test-Retest Relability of the Torrance Creativity Battery

Test-retest reliabilities were obtained for the creativity tests by correlating the pre- and post-test creativity test scores. The reader will recall that the pre-test was administered in the fall and the post-test in the spring. The interval between testings was approximately six months, and an alternate form of the creativity test battery was administered in the spring with the exception of the Circles Test for which no alternate form was available. The reliability coefficients were computed separately within each of the four grades. The reliability coefficients for the non-verbal creativity tests and the non-verbal battery total scores are shown in Table 6.11 for grades two through five. The reliabilities obtained for the scores derived from the three tests in the non-verbal battery (e.g., Picture Construction, Figure Completion, and Circles) ranged from -.06 to .62. The Circles Test was clearly the most reliable of the three non-verbal tests having reliabilities ranging from .30 to .62. There was also a tendency for the reliability coefficients to get higher as grade level increased. The reliabilites were generally higher in grades four and five than in grades two and three. The reliability of the total non-verbal score was .46 in grade two, .34 in grades three, .61 in grade four, and .64 in grade five.

The higher reliabilities obtained for the Circles Test may have resulted from the fact that identical forms of the test were used in the pre- and post-test administrations. Little data was available on the equivalence of the alternate forms in the Torrance creativity battery. The lower reliabilities of the Picture Construction and Figure Completion subtests may have been due to the lack of equivalence of the alternate forms. Although the alternate forms appear to call for similar responses on the part of subjects, the possibility exists that minor differences in the stimulus materials produced differences in the kinds of responses produced by the pupils.

The higher reliabilities of the Circles Test may also have resulted from the fact that the Circles Test is the longest of the three non-verbal subtests in terms of the number of responses produced by pupils. Only one drawing is scored in the Picture Construction Test, a maximum of six drawings may be scored in the Figure Completion Test, and a maximum of 42 drawings may be produced in the Circles Test. Since each response in these creativity tests is somewhat analogous to an item in a traditional psychological test, the increased reliablity of the Circles Test may have been a function of the increased number of items responded to on the test.

The reliabilities of the verbal creativity subtests and total scores for grades four and five are shown in Table 6.12. Generally higher reliabilities were obtained for verbal creativity subtests as compared to non-verbal subtests. The reliability coefficients of the verbal creativity subtest scores ranged from .21 to .74, while those for verbal battery totals ranged from .35 to .79. The most reliable score derived from the battery was the Verbal Fluency total score having a reliability of .79.

As was the case for the non-verbal creativity scores, the verbal scores were more reliable in the upper grade. This finding was consistent with the results obtained for more traditional intelligence tests, although the increase in the reliability of intelligence tests typically occurs much earlier. Measures of intelligence at the pre-school level typically have relatively low reliability, whereas measures of intelligence employed when children enter school reach acceptable levels of reliability. This tendency for reliabilities to increase with age may be partly a function of the difficulties involved in testing younger

TABLE 6.11

NON-VERBAL CREATIVITY TEST-RETEST RELIABILITIES FOR GRADES TWO THROUGH FIVE*
(Interval between testings = six months)

CREATIVITY TEST SCORE	RELIABILITIES			
	GRADE 2	GRADE 3	GRADE 4	GRADE 5
Picture Construction:				
Originality	-.04	.06	.04	.03
Elaboration	.13	.04	.27	.11
Activity	.08	.12	.26	.04
Title Adequacy	.36	.21	.25	.22
Total	.21	.01	.31	.05
Figure Completion:				
Fluency	-.06	.23	.14	.35
Flexibility	.07	-.01	.09	.18
Originality	.13	.07	.17	.20
Elaboration	.13	.10	.15	.36
Title Adequacy	.12	.31	.29	.30
Total	.05	.16	.17	.35
Circles:				
Fluency	.52	.34	.60	.61
Flexibility	.46	.30	.32	.37
Originality	.48	.35	.60	.55
Elaboration	.49	.34	.62	.62
Total	.51	.36	.61	.59
Non-Verbal Battery Totals:				
Fluency	.49	.35	.56	.63
Flexibility	.40	.34	.33	.40
Originality	.41	.28	.59	.59
Elaboration	.40	.29	.53	.45
Title Adequacy	.23	.38	.33	.38
Total Non-Verbal	.46	.34	.61	.64

*N Grade 2 = 158-183
N Grade 3 = 111-123
N Grade 4 = 105-120
N Grade 5 = 132-147

TABLE 6.12

VERBAL CREATIVITY TEST-RETEST RELIABILITIES FOR GRADES FOUR AND FIVE*

(Six-month interval between testings)

CREATIVITY TEST SCORE	RELIABILITIES	
	GRADE 4	GRADE 5
Ask and Guess Test:		
Fluency	.53	.74
Adequacy	.44	.57
Flexibility	.30	.49
Total	.50	.66
Product Improvement:		
Fluency	.41	.49
Flexibility	.28	.29
Originality	.31	.47
Elaboration	.38	.51
Total	.36	.50
Unusual Uses:		
Fluency	.28	.38
Flexibility	.21	.40
Originality	.25	.38
Elaboration	.30	.44
Total	.28	.43
Consequences:		
Fluency	.42	.68
Flexibility	.39	.32
Total	.41	.58
Verbal Battery Totals:		
Fluency	.56	.79
Flexibility	.50	.61
Originality	.35	.58
Elaboration	.41	.61
Total Verbal	.53	.73
Total Creativity	.61	.75

*N Grade 4 = 105-120
N Grade 5 = 132-147

children and partly a result of the gradual emergence of the trait being measured as a stable trait.

The reliabilities obtained in the present investigation were similar to those reported by Torrance for these tests. Torrance reported reliabilities ranging from the upper .40's to the upper .60's for battery total scores when the subjects used were elementary school pupils. Higher reliabilities were reported by Torrance for a sample of college students ranging from .68 to .85 for subtests, and .88 for battery total scores. A three-month interval was employed between testings in the Torrance Study.

Relationships Between Measures of Creativity and Intelligence

The results of the present investigation were consistent with the results of other studies concerning the question of the relationship between measures of intelligence and creativity. Zero order correlations between intelligence and creativity measures have usually been obtained when the researcher has already selected his subjects on the basis of high intellectual ability or high achievement. In cases where correlations were obtained for unrestricted samples or when the data were corrected for restricted range, correlations ranging from .20 to .40 were the rule. The present research employed an unrestricted sample of elementary school pupils as subjects. The correlations between non-verbal creativity, verbal creativity, and non-verbal and verbal intelligence as measured by the Lorge-Thorndike are reported in Table 6.13. The correlations between non-verbal creativity test scores, and verbal, non-verbal, and total intelligence scores ranged from -.07 to .33 for grades two through five with a median r of .23. The correlations between total non-verbal creativity and non-verbal intelligence were .27, .30, .14, and .29 for grades two through five respectively, suggesting that the relationship between creativity and intelligence was not entirely the result of a verbal ability factor. The correlations between verbal creativity scores and intelligence for the fourth and fifth grades were consistently higher than those for non-verbal creativity ranging from .27 to .52 with a median r of .43. Other studies have reported a lowering of correlations between verbal creativity and intelligence in the upper elementary school grades, presumably the result of the decreasing importance of writing ability in the group administration of creativity tests. The beginnings of such a trend of lowering correlations from the fourth to the fifth grade was not observed in the present study. The correlations between verbal creativity and intelligence were very similar in the fourth and fifth grades. It is possible that the aforementioned trend would have occurred if grades beyond the fifth grade had been tested.

TABLE 6.13

RELATIONSHIPS BETWEEN INTELLIGENCE AND CREATIVITY IN GRADES TWO THROUGH FIVE*

CREATIVITY TEST SCORE	INTELLIGENCE							
	GRADE 2	GRADE 3	GRADE 4			GRADE 5		
	N.V.	N.V.	V.	N.V.	T.	V.	N.V.	T.
Non-Verbal Battery Totals:								
Fluency	.24	.26	.23	.11	.18	.30	.27	.31
Flexibility	.20	.14	.29	.10	.21	.24	.30	.29
Originality	.25	.33	.16	.10	.15	.30	.29	.32
Elaboration	.21	.16	.25	.20	.25	.23	.21	.23
Title Adequacy	.23	.09	.04	-.07	-.01	.20	.19	.21
Total Non-Verbal	.27	.30	.23	.14	.21	.31	.29	.32
Verbal Battery Totals:								
Fluency			.42	.34	.41	.46	.34	.43
Flexibility			.45	.43	.47	.46	.34	.43
Originality			.39	.27	.36	.50	.37	.47
Elaboration			.45	.36	.44	.48	.38	.46
Total Verbal			.48	.39	.47	.52	.40	.49
Total Creativity			.45	.36	.44	.46	.37	.45

*Grade 2 N = 158-183 $r_{.05} = .16$

Grade 3 N = 111-123 $r_{.05} = .19$

Grade 4 N = 105-120 $r_{.05} = .20$

Grade 5 N = 132-147 $r_{.05} = .18$

CHAPTER VII

CONCLUSIONS

During the past four years considerable data have been collected on sixty-five teachers and their students in grades one through five. Teacher measures include ratings and Q-sorts obtained after classroom observation; scores based on verbatim recording of a sample of verbal statements in the classroom as well as scores based on the OSCAR device and scores on several paper and pencil devices of both a projective and questionnaire nature. Pupil data include regressed gain scores in the areas of reading vocabulary, reading comprehension, and arithmetic; questionnaire measures of test anxiety, dependency, achievement and affiliation needs, liking for school, perception of the teacher as controlling, achievement oriented and affiliative, and scores on three of the creativity tests developed by Torrance. The Russell Sage Test of Social Relations was administered in grades two through five.

The sixty-five teachers are a sub-sample (selective only in terms of permanence in the school system) from a larger sample of 120 teachers utilized in a previously reported study which demonstrated that the extent of controlling behavior within the classroom was predictable from a paper and pencil questionnaire (Wallen and others, 1963). Six elementary schools are represented in this sample and the distribution across grades yields approximately twelve teachers in each of grades one through five.

For an additional sub-sample consisting of nineteen teachers in grades two through five selected at the extremes of the controlling dimension, additional pupil data were collected. These data consisted of gain scores on the Torrance creativity tests and measures of the classroom behavior of students selected on the basis of intelligence and creativity tests.

Analysis of the data consisted of correlational analyses and factorial analyses wherein selected combinations of teacher and student characteristics were used as main effects and other student measures constituted the dependent variable.

It is clear that any conclusions arrived at as a result of a study such as this must be held quite tentatively. The first reason for this is that our sampling of teachers does not constitute a sample drawn randomly from teachers in general or even from teachers within the particular locale in which the study was conducted. We do have some confidence in generalizing results, at least within this geographical area, due to our attempt to obtain teachers from a number of schools and our evidence (from our earlier study) that the more important results pertaining to the prediction of teacher behavior were replicated with both sub-samples. Nevertheless, this caution should be kept in mind.

Secondly, most of the hypotheses which we originally developed and which were directly tested were not supported. Thus, we have in large measure resorted to an empirical examination of the many relationships possible within our matrix of variables. Although we have attempted to guard against capitalization on chance, there remains the possibility that some of our findings reflect nothing more than chance fluctuations. This problem is particularly evident when the relationships are examined within a particular grade with the smaller N which this necessitates. Unfortunately, within-grade analysis became imperative in this study due to the finding that one of our initial assumptions, i.e., that any relationships obtained between teacher characteristics and student characteristics would take essentially the same form across all five grades is, in many instances, apparently an incorrect assumption, although one cannot be sure whether it is the relationship between the fundamental variables themselves which differs or whether certain aspects of the measurement procedures may have differed from grade to grade.

Nevertheless, it is possible to place our findings along a continuum of tentativeness and we shall first list those results which seem to us to have considerable support within our data, both in terms of consistencies across the five grades and in terms of the consistency of relationships among the variables themselves.

1. Achievement gain in Reading Vocabulary appears to be positively correlated with the extent to which the teacher is viewed as a stimulating teacher by observers. This relationships seems to exist within all five grades.

2. Liking for School appears to be related to a teacher orientation which is less achievement oriented as viewed by observers. It is probably to be expected that the teacher who makes the greatest achievement and academic demands on the student will engender the greater amount of frustration on the part of the poor students in particular, and hence, a somewhat less favorable attitude toward her and school in general.

3. Liking for School seems to be positively related to the degree of warmth and permissiveness of the teacher in the upper grades. In the lower grades, and particularly in grade 1, this relationship is possibly reversed, although this finding is considered very tentative due to the complications in interpretation of the measure of permissiveness encountered in this grade.

4. The extent to which the teacher is viewed by observers as achievement oriented seems to have a positive affect upon the ability of the class to plan effectively in a group problem solving task, but also appears to foster a breakdown in group relationships in the actual operations phase of the problem through this did not appear until the second of two problems was encountered.

5. Supportive behavior on the part of the teacher appears to foster a more friendly group interaction during the operations phase of group problem solving activity.

6. Supportive behavior on the part of the teacher appears to be negatively correlated with the extent of test anxiety within all five grades.

7. There appears to be a fair degree of correspondence between observer and pupil perception of the teacher in terms of extent of affiliation behavior beginning at about the third grade. In the first two grades this correspondence is extremely poor.

8. Speaking only of achievement gain in the three areas tested, it appears that some teachers do a better job of teaching reading, across all grades than others. Within the first grade, there is the suggestion that the teacher who fosters greater gain in Reading Comprehension also fosters greater gain in Arithmetic. This trend was also observed in the upper grades but may (in the upper grades) be attributable to differences in class intelligence.

9. Coefficients of stability for the Torrance creativity tests over a six-month period show greater stability in the higher grades and higher stability for the verbal than for the non-verbal measures. In grades four and five, the 'total' score in each area (verbal and non-verbal) provides stability coefficients of the magnitude of .60 to .75. In the lower grades (two and three) the non-verbal 'total' stability values are .46 and .34.

Next we may list a series of findings which we regard as even more tentative than those listed above.

1. It appears that the teacher who is excessively affiliative or affectionate in the first grade and who, in addition, has a strong need for affiliation is likely to be less well liked and to engender less of an achievement gain than the less affiliation oriented teacher, although hostility in the first grade on the part of the teacher appears undesirable. It appears that verbal supportive behavior, praise, and encouragement on the part of the teacher is quite important in the first grade, fostering both liking for school and gain in achievement. It also appears, however, that in the first grade the student may have need of considerable structure within the school environment such that the more controlled classroom and the controlling, somewhat aloof teacher may provide an environment within this grade which is more comfortable and more conducive to achievement for the first grade student. It would appear that the first grade teacher should provide a very well ordered, well organized classroom which does not permit much pupil decision but which is at the same time comforting and encouraging but not overtly affectionate.

2. There is the suggestion of an interaction between the teacher characteristic of control and the pupil characteristic of dependency in that within the first grade on the reading vocabulary measure the children who were high in dependency need, as measured by our questionnaire, did better with the more controlling teachers, whereas, the children who were low in dependency need did better with the less controlling teachers.

One analysis suggests the same effect to be true for arithmetic gain. There is the further suggestion that this relationship is reversed by the fifth grade in that the less dependent children achieved less achievement gain in arithmetic with the less controlling teachers.

3. With respect to changes in creativity, at least as measured by the Torrance tests, there is some evidence that the more permissive teacher in the upper grades tends to foster a greater change in the direction of more originality of thinking than does the more controlling teacher.

4. There is the suggestion that pupils, at least in the upper grades, initiate more verbal exchange with the less controlling teacher.

5. On several fairly reliable indexes of pupil classroom behavior there was no indication of differing behavior on the part of 'high' and low 'creative' pupils.

In addition to the generalizations just discussed, it seems worth recording the overall impression which we are left with; an impression which, it should be recognized, would not necessarily be shared by other individuals working with the same data since it does not have clear-cut support from the data. Nevertheless, the impression is one of the importance of a developmental point of view with regard to teacher behavior acorss the first five grades. We have the distinct impression from our data that the typical first grader, being somewhat unsure of himself and new to the situation, is both more comfortable and achieves better given a situation which is quite structured, quite controlled by the teacher while at the same time being rather supportive and encouraging but without overt affection on the part of the teacher. Data which tend to support this are the preceding discussion of the results within grade one as well as the finding that this type of description seems to hold particularly true for the more dependent child. In about the second grade it is our impression that this pattern tends to change and that by the upper grades the general desirability of encouragement still exists but the effect of control shifts such that a greater degree of permissiveness has the more desirable effects in terms of both liking for school and achievement gain.

Further, the totaility of results strongly suggest that relationships between teacher characteristics and pupil behavior may be very different in the first one or two grades than for the succeeding grades. Since almost all research has been done using grades four and above (probably largely for reasons of convenience which we can well appreciate), we suggest that: (1) generalization below these grades is very questionable and (2) studies in the first grade seem imperative.

An additional grant from the U. S. Office of Education will permit us to further analyze these data. At the present time we contemplate several approaches as follows.

1. We hope to obtain additional evidence pertaining to our interpretation of the results in the first grade. This will consist in part

of analyzing another measure obtained in our earlier study which consists of ratings of pictures drawn by the students of their class and teacher. A re-classification of our Verbal Statements measure which eliminates the reprimanding category from the control score should prove illuminating in this regard.

2. We intend to conduct further factorial analyses employing different combinations of teacher and pupil characteristics. As a part of this we may divide our sample of teachers along the same type of dimension which seems to have been profitable in the study by Heil (1960).

3. A factor analysis of the pupil matrix within each grade should prove of interest since we have on hand a large number of measures on these pupils.

REFERENCES

REFERENCES

Boneau, C. A. The effects of violations of assumptions underlying the t-test. *Psychol. Bull.*, 1960, *57*, 49-64.

Damrin, D. E. *Manual for the Russell Sage Social Relations Test.* Princeton: Educational Testing Service, 1959.

Damrin, D. E. The Russell Sage Social Relations Test: A technique for measuring group problem solving skills in elementary school children. *J. exp. Educ.*, 1959, *28*, 85-99.

Christensen, C. M. Relationships between pupil achievement, pupil affect-need, teacher warmth, and teacher permissiveness. *J. educ. Psychol.*, 1960, *51*, 169-174.

Cogan, M. L. The behavior of teachers and the productive behavior of their pupils: I. "Perception" analysis. *J. exp. Educ.*, 1958, *27*, 89-105. (a)

Cogan, M. L. The behavior of teachers and the productive behavior of their pupils: II. "Trait" analysis. *J. exp. Educ.*, 1958, *27*, 106-124. (b)

Cox. F. N. Correlates of general and test anxiety in children. *Australian J. Psychol.*, 1960, *12*, 169-177.

Flanders, N. A. *Teacher Influence, Pupil Attitudes and Achievement.* Pre-publication manuscript of proposed research monograph for U. S. Office of Education, Cooperative Research Branch, Research Project No. 397, 1962.

Heil, L. M., Powell, Marion, and Feifer, I. *Characteristics of Teacher Behavior Related to the Achievement of Children in Several Elementary Grades.* Final Report, U. S. Office of Education, Cooperative Research Branch, Research Project No. SAE 7285, 1960.

Medley, D. M. and Klein, A. A. Measuring classroom behavior with a pupil reaction inventory. *Elem. sch. J.*, 1957, *57*, 315-319.

Medley, D. M. and Mitzel, H. E. A technique for measuring classroom behavior. (Manual for OSCAR) Division of Teacher Education of the Municipal Colleges of New York City. Mimeographed, undated.

Medley, D. M. and Mitzel, H. E. A technique for measuring classroom behavior. *J. educ. Psychol.*, 1958, *49*, 86-92. (a)

Medley, D. M. and Mitzel, H. E. Application of analysis of variance to the estimation of the reliability of observations of teachers' classroom behavior. *J. exp. Educ.*, 1958, *27*, 23-25. (b)

Medley, D. M. and Mitzel, H. E. Some behavioral correlates of teacher effectiveness. J. educ. Psychol., 1959, 50, 239-246.

Medley, D. M. and Mitzel, H. E. Measuring classroom behavior by systematic observation. In Gage, N. L. (Ed.) Handbook of Research on Teaching, Chicago: Rand McNally, 1963, 247-328.

Mitzel, H. E. Teacher Effectiveness. In C. W. Harris (Ed.) Encyclopedia of Educational Research, New York: Mac Millan, 1960, 1482-1486.

Mitzel, H. E. and Medley, D. M. Pupil growth in reading--an index of effective teaching. J. educ. Psychol., 1957, 48, 227-239.

Mitzel, H. E. and Rabinowitz, W. Assessing social-emotional climate in the classroom by Withall's technique. Psychol. Monogr., 1953, 67, 18, (Whole No. 368).

Phillips, B. H. Sex, social class, and anxiety as sources of variation in school achievement. J. educ. Psychol., 1962, 53, 316-322.

Rabinowitz, W. and Travers, R. M. W. Problems of defining and assessing teacher effectiveness. Ed. Theory, 1953, 3, 212-219.

Remmers, H. H. (Ch.) Report of the Committee on the criteria of teacher effectiveness. Rev. educ. Res., 1952, 22, 238-263.

Remmers, H. H. (Ch.) Second report of the Committee on criteria of teacher effectiveness. J. educ. Res., 1953, 46, 641-658.

Ryans, D. G. Some validity extension data from empirically derived predictors of teacher behavior. Educ. psychol. Meas., 1958, 18, 355-370.

Ryans, D. G. Characteristics of Teachers. Washington, D. C.: American Council on Education, 1960.

Sarason, S. B., et al. Anxiety in Elementary School Children. New York: Wiley and Sons, 1960.

Silberman, E. F. Effects of praise and reproof on reading growth in a non-laboratory classroom setting. J. educ. Psychol., 1957, 48, 199-206.

Stern, G. G. and Masling, J. M. Unconscious factors in career motivation for teaching. Final Report, U. S. Office of Education, Cooperative Research Branch, Research Project No. SAE 6459, 1958.

Torrance, E. P., Yamamoto, K., Schnetzki, D., Palamutlu, N., and Luther, B. Assessing the creative abilities of children. Bureau of Educational Research, College of Education, University of Minnesota, 1960.

Travers, R. M. W., Wallen, N. E., Reid, I. E., and Wodtke, K. H. Measured needs of teachers and their behavior in the classroom. Salt Lake City: University of Utah, 1961. (Mimeographed)

Wallen, N. E. and Stevenson, G. M. Stability and correlates of judged creativity in writing. J. educ. Psychol., 1960, 51, 273-276.

Wallen, N. E. and Travers, R. M. W. Analysis and investigation of teaching methods. In Gage, N. L. (Ed.) Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 448-505.

Wallen, N. E., Travers, R. M. W., Reid, I. E., and Wodtke, K. H. Relationships between teacher needs and teacher behavior in the classroom. J. educ. Psychol., 1963, 54, 23-32.

Wandt, E. A comparison of the attitudes of contrasting groups of teachers. Educ. psychol. Meas., 1954, 14, 418-422.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Withall, J. Development of a technique for the measurement of socio-emotional climate in classrooms. J. exp. Educ., 1949, 17, 347-361.

Wodtke, K. H., Reid, I. E., Wallen, N. E., and Travers, R. M. W. Patterns of needs as predictors of classroom behavior of teachers. Educ. psychol. Meas. In Press.

Yamamoto, K. Scoring Manual for Evaluating Imaginative Stories. Minneapolis: Bureau of Educational Research, University of Minnesota, 1961.

Yamamoto, K. Revised Scoring Manual for Tests of Creative Thinking. Minneapolis: Bureau of Educational Research, University of Minnesota, 1962.

APPENDIX A

SCORING ADDITIONS TO RUSSELL SAGE TEST OF SOCIAL RELATIONS

SCORING ADDITIONS TO RUSSELL SAGE TEST OF SOCIAL RELATIONS

Planning Stage:

A. Participation.

4. Fewer than 10 checks at or above organized groups with little or no examiner action.

3. Ten or more checks at or above unorganized groups with marked examiner interference.

B. Communication.

4. The discussion has fewer than 8 participants but centers above organized groups. Fifty percent or more of checks are evaluating improving.

3. Shofter discussion at or above unorganized groups.

F. Autonomy.

4S. Lengthy discussion with scoeral prods (3 or more)--good communication pattern--good involvement.

3S. Shorter discussion (with prods) but involvement must be ++ or +- (5 to 7).

2R. Shorter discussion (with reminds), involvement ++ or +-.

Operations Stage:

B. Atmosphere.

2b Quiet--Noisy.

5. Checked Friendly or Quiet throughout.

4. Only an occasional check for noisy or excited.

3. Builds to noisy or excited in middle stage and remains.

2. Noise and excited throughout or builds gradually to boisterou screaming in end period.

1. Checked for boisterous or screaming in all three periods.

2c **Friendly--Hostile.**

5. Checked Friendly throughout.

4. Only an occasional check under bickering or quiet, suppressed, or a few checks under excited.

3. Builds to bickering in the middle stage and remains or checks under tense and excited with an occasional bicker.

2. Builds quickly to bickering--occasional check under angry, quarrelsome.

1. Bickering and angry quarrelsome throughout.

D. Success.

3. Or, less than eleven minutes with many errors.

2. Or, completed in 11 to 15 minutes but with many errors.

APPENDIX B

BETAS, MEANS, STANDARD DEVIATIONS AND MULTIPLE r's FOR ACHIEVEMENT GAIN SCORES--BY GRADE

Note 1.--Variable 1 is CMM. Cariable 2 is the pre-test and variable 3 the post-test or predicted variable.

Note 2.--Differences among grades on means and standard deviations are, in part, due to differing levels of the tests.

Note 3.--Slight differences in CMM values for mean and S.D. reflect minor changes in students constituting the sample due to absenteeism.

READING VOCABULARY

GRADE	β_1	β_2	$\bar{X}_1$	s_1	$\bar{X}_2$	s_2	$\bar{X}_3$	s_3	$r_{12.3}$
1	.31	.41	53.9	10.5	22.7	11.2	51.0	11.3	.64
2	.21	.59	66.4	7.9	56.1	10.8	66.8	6.9	.80
3	.19	.64	74.1	6.5	34.6	6.7	25.6	7.7	.90
4	.24	.58	65.9	15.2	28.2	9.3	34.1	7.3	.77
5	.40	.47	79.4	16.2	34.9	9.0	39.1	7.2	.81

READING COMPREHENSION

GRADE	β_1	β_2	$\overline{X}_1$	s_1	$\overline{X}_2$	s_2	$\overline{X}_3$	s_3	$r_{12.3}$
1	.39	.00	53.9	10.5	0.1	1.8	5.6	3.9	.39
2	.26	.42	66.4	7.9	6.8	4.1	12.4	2.9	.59
3	.11	.70	74.1	6.5	34.6	9.7	31.1	11.4	.76
4	.15	.69	65.9	15.2	36.7	12.0	43.8	12.8	.81
5.	.36	.53	79.4	16.2	46.1	11.8	52.1	11.2	.84

ARITHMETIC

GRADE	β_1	β_2	X_1	s_1	X_2	s_2	X_3	s_3	$r_{12.3}$
1	.25	.30	53.9	10.5	6.5	8.2	27.9	10.7	.46
2	.18	.48	66.4	7.9	31.2	8.0	40.5	4.3	.60
3	.22	.35	74.1	6.6	94.6	21.1	19.2	5.9	.50
4	.34	.31	67.4	14.6	23.4	6.4	29.9	6.6	.55
5	.39	.42	81.7	16.2	33.3	8.0	40.9	10.6	.74

APPENDIX C

OBSERVATION RECORD AND GLOSSARY OF RATINGS FOR PUPIL BEHAVIOR

C.02

Pupil __________________ Teacher ________________ Observer ___________

Q.	"How to do it"											
O.	"Can I do it"											
Q.	"Why"											
A.	Rds, stds at seat											
A.	Writes at seat											
A.	Works at board											
A.	Recites											
A.	Reads aloud											
M.	Gms or exer.											
M.	Gleans board											
M.	Rests, has snk.											
M.	Lvs. enters rm.											
M.	Cllcts, psses											
L.	Chosen to lead											
L.	Leads spont.											
NC	Fls to fllw orders											
NC	"Goofs off"											
NC	Reprmd by T											
NC	Igns T's ques.											
NC	Scuffles, fights											
NC	Whispers, laughs											
NC	Boisterous, scream.											
NC	Angry, quarrelsome											
NC	Bicker, squabble.											
NC	Noisy, joking											
PT	Responds to T											
PT	Self-init. talk											
G.	P works alone											
G.	P works in group											
I.	T/P init. by P											
I.	T/P init. by T											
E.	Pos affect twds T											
E.	Neg affect twds T											
E.	Neg affect twds P's											
Id	Suggts diff, new idea											
Id	Rpts idea											
Id	Eval. imprvs idea											
P	Init. spont act acad gets bk, tks from desk											
P	Init. spont act non-acad gets drnk, lvs rm											

Ratings:

Apathetic 1 2 3 4 5 6 7 N Alert

Obstructive 1 2 3 4 5 6 7 N Responsible

Uncertain 1 2 3 4 5 6 7 N Confident

Dependent 1 2 3 4 5 6 7 N Initiating

GLOSSARY FOR RATINGS*

APATHETIC

1. Listless.
2. Bored-acting.
3. Entered into activities half-heartedly.
4. Restless.
5. Attention wandered.
6. Slow in getting under way.

OBSTRUCTIVE

1. Rude to one another and/or to teacher.
2. Interrupting; demanding attention; disturbing.
3. Obstinate; sullen.
4. Refusal to participate.
5. Quarrelsome; irritable.
6. Engaged in name calling and/or tattling.
7. Unprepared.

UNCERTAIN

1. Seemed afraid to try; unsure.
2. Hesitant; restrained.
3. Appeared embarrassed.
4. Frequent display of nervous habits, nail-biting, etc.
5. Appeared shy and timid.
6. Hesitant and/or stammering speech.

DEPENDENT

1. Relied on teacher for explicit directions
2. Showed little ability to work things out himself.
3. Unable to proceed when initiative called for.
4. Appeared reluctant to take lead or accept responsibility

ALERT

1. Appeared anxious to recite and participate
2. Watched teacher attentively.
3. Worked concentratedly.
4. Seemed to respond eagerly.
5. Prompt and ready to take part in activities when they begin.

RESPONSIBLE

1. Courteous, cooperative, friendly with each other and with teacher.
2. Completed assignments without complaining or unhappiness.
3. Controlled voices.
4. Received help and criticism attentively.
5. Asked for help when needed.
6. Orderly without specific directions from teacher.
7. Prepared.

CONFIDENT

1. Seemed anxious to try new problems or activities.
2. Undisturbed by mistakes.
3. Volunteered to recite.
4. Entered freely into activities.
5. Appeared relaxed.
6. Spoke with assurance.

INITIATING

1. Volunteered ideas and suggestions.
2. Showed resourcefulness.
3. Took lead willingly.
4. Assumed responsibilities without evasion.

---

* From Ryans (1960)

APPENDIX D

PUPIL QUESTIONNAIRE--KEYED

Note.--Items 1-30 comprise the Sarason Test Anxiety Scale.
Items 76-80 comprise five of the eight items used by Medley and Klein to appriase 'liking for the teacher.'

The remaining items comprise six scales of each constructed for this project and intended to reflect three pupil needs: achievement (NAch), affiliation (NA), and dependency (ND) (10 items each) and perception of the teacher as: achievement oriented (PAch), affiliative (PA), and controlling (PC) (5 items each). Items for each of these are identified.

Each item of the Anxiety Scale is scored one point for a yes response. The remaining items are keyed for the response (yes--Y or no--N) which increases the score by 1 point.

PUPIL QUESTIONNAIRE

1. Do you worry when the teacher says that she is going to ask you questions to find out how much you know?

2. Do you worry about being promoted, that is, passing from the _____ to the _____ grade at the end of the year?

3. When the teacher asks you to get up in front of the class and read aloud, are you afraid that you are going to make some bad mistakes?

4. When the teacher says that she is going to call upon some boys and girls in the class to do arithmetic problems, do you hope that she will call upon someone else and not on you?

5. Do you sometimes dream at night that you are in school and cannot answer the teacher's questions?

6. When the teacher says that she is going to find out how much you have learned, does your heart begin to beat faster?

7. When the teacher is teaching you about arithmetic, do you feel that other children in the class understand her better than you?

8. When you are in bed at night, do you sometimes worry about how you are going to do in class the next day?

9. When the teacher asks you to write on the blackboard in front of the class, does the hand you write with sometimes shake a little?

10. When the teacher is teaching you about reading, do you feel that other children in class understand her better than you?

11. Do you think you worry more about school than other children?

12. When you are at home and you are thinking about your arithmetic lesson for the next day, do you become afraid that you will get the answers wrong when the teacher calls upon you?

13. If you are sick and miss school, do you worry that you will do more poorly in your schoolwork than other children when you return to school?

14. Do you sometimes dream at night that other boys and girls in your class can do things you cannot do?

15. When you are home and you are thinking about your reading lesson for the next day, do you worry that you will do poorly on the lesson?

16. When the teacher says that she is going to find out how much you have learned, do you get a funny feeling in your stomach?

17. If you did very poorly when the teacher called on you, would you probably feel like crying even though you would try not to cry?

18. Do you sometimes dream at night that the teacher is angry because you do not know your lessons?

In the following questions the word "test" is used. What I mean by "test" is any time the teacher asks you to do something to find out how much you know or how much you have learned. It could be by your writing on paper, or by your speaking aloud, or by your writing on the blackboard. Do you understand what I mean by "test"--it is any time the teacher asks you to do something to find out how much you know.

19. Are you afraid of school tests?

20. Do you worry a lot before you take a test?

21. Do you worry a lot while you are taking a test?

22. After you have taken a test do you worry about how well you did on the test?

23. Do you sometimes dream at night that you did poorly on a test you had in school that day?

24. When you are taking a test, does the hand you write with shake a little?

25. When the teacher says that she is going to give the class a test, do you become afraid that you will do poorly?

26. When you are taking a hard test, do you forget some things you knew very well before you started taking the test?

27. Do you wish a lot of times that you didn't worry so much about tests?

28. When the teacher says that she is going to give the class a test, do you get a nervous or funny feeling?

29. While you are taking a test do you usually think you are going to do poorly?

30. While you are on your way to school, do you sometimes worry that the teacher may give the class a test?

ND 31. Do you like to have your teacher tell you exactly how to do your work? Y

NA 32. Do you want everyone to like you? Y

NAch 33. Do you like to read outside of school? Y

NAch 34. Do you like doing your arithmetic? Y

NA 35. Would you give up your recess to help a friend? Y

NA 36. Do you like to be around lots of people? Y

ND 37. Do you like to have to figure out new ways to do things? N

NA 38. Do you sometimes feel that you don't care what other people think of you? N

NAch 39. If there is something you don't know, do you like to find out about it yourself? Y

NA 40. Do you like to learn new things even if it's hard? Y

NAch 41. Do you like to learn new things even if it's hard? Y

NAch 42. Would you work during recess to get a better mark? Y

ND 43. Do you believe everything your teacher tells you? Y

NA 44. Does it bother you if other children have more friends than you do? Y

ND 45. Do you think children should always do what their teacher tells them to do? Y

ND 46. Do you think children should ever disagree with their parents? N

NA 47. Are you unhappy if you have to walk home from school alone? Y

NAch 48. Do you like to take tests to show how much you know? Y

ND 49. When you play games would you rather someone else was the leader instead of yourself? Y

ND 50. Do you like to feel that your teacher will always take care of you? Y

NAch 51. Are you usually satisfied with your school work? N

NAch 52. Do you like to ask questions? Y

NA 53. Does it make you feel bad to get angry at someone else? Y

ND 54. Do you like the kind of teacher who will always help you do things? Y

NAch 55. Do you like to do extra work? Y

NAch 56. Do you often wonder why things happen the way they do? Y

ND 57. Do you like to have your teacher help you with your work? Y

NA 58. Do you like to have grown ups, besides your parents, call you dear? Y

NA 59. Do you like to do things with your classmates rather than by yourself? Y

ND 60. Do you like to tell your troubles to your teacher? Y

PA 61. Do you think of your teacher as a friend? Y

PC 62. Do you have to do lots of things in school that you don't want to do? Y

PC 63. In school are you always told what to do and when to do it? Y

PAch 64. Does your teacher expect everyone to do their very best? Y

PC 65. Do you help plan what the class is going to do? N

PA 66. Would you like to be with your teacher outside of school? Y

PA 67. Is your teacher ever mean? N

PAch 68. Do you always find out whether your work is right? Y

PC 69. Can you leave your seat without asking? N

PA 70. Do you hope all your teachers are like the one you have now? Y

PC 71. Do children ask a lot of questions in class? N

PAch 72. Is your teacher mostly interested in how much you learn? Y

PAch 73. Is your teacher very often wrong? N

PA 74. Is your teacher mostly interested in whether you are happy? Y

PAch 75. Do you think you have learned a lot from your teacher? Y

L 76. Do you ever feel like staying away from school? N

L 77. Do you have much fun in this class? Y

L 78. Do you always do your best in this class? Y

L 79. Do most of the children like the teacher? Y

L 80. Does the teacher help enough? Y